UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE DIREITO
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO E INOVAÇÃO

JESSICKA OLIVEIRA DE ASSIS

Colonialismo digital e ditadura de dados: o papel da accountability algorítmica sobre agentes inteligentes na democracia moderna

Juiz de Fora
2024

Jessicka Oliveira de Assis

Colonialismo digital e ditadura de dados: o papel da accountability algorítmica sobre agentes inteligentes na democracia moderna

Dissertação apresentada à Faculdade de Direito da Universidade Federal de Juiz de Fora para obtenção do título de Mestre em Direito pelo Programa de Pós-graduação em Direito e Inovação.

Área de concentração: Direito e inovação

Linha II - Direitos Humanos, Pessoa e Desenvolvimento: inovação e regulação jurídica no contexto do capitalismo globalizado

Orientador: Professor Doutor Wagner Silveira Rezende

Juiz de Fora

2024

Oliveira de Assis, Jessicka.
Colonialismo digital e ditadura de dados: o papel da accountability algorítmica sobre agentes inteligentes na democracia moderna / Jessicka Oliveira de Assis. -- 2024.
135 f.

Orientador: Wagner Silveira Rezende
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Direito. Programa de Pós-Graduação em Direito, 2023.

1. Colonialismo digital e a ditadura dos dados 2. Estado da arte no cenário jurídico 3. Metodologia 4. Recorte dos princípios do arcabouço legal constituído 5. Análise qualitativa do recorte constituído. Silveira Rezende, Wagner, orient. II. Título.

JÉSSICKA OLIVEIRA DE ASSIS

Colonialismo digital e ditadura de dados: o papel da accountability algorítmica sobre agentes inteligentes na democracia moderna

Dissertação apresentada ao Programa de Mestrado em
Direito da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do titulo de Mestre em Direito. Área de concentração: Direito e Inovação

Aprovada em 30 de abril de 2024.

BANCA EXAMINADORA

Wagner Silveira
Rezende -
Orientador
Universidade
Federal de Juiz de
Fora

Cora Hisae
Monteiro da Silva
Hagino
Universidade
Federal Fluminese

Clarissa Diniz Guedes

Universidade Federal
de Juiz de Fora

Caroline Pinheiro
da Rosa
Universidade
Federal de Juiz de
Fora

Juiz de Fora, 24/04/2024.

Documento assinado eletronicamente por **Wagner Silveira Rezende**, **Professor(a)**, em 21/05/2024, às 09:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **CORA HISAE MONTEIRO DA SILVA HAGINO**, **Usuário Externo**, em 29/05/2024, às 16:38, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jessicka Oliveira de Assis**, **Usuário Externo**, em 16/07/2024, às 16:07, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Clarissa Diniz Guedes**, **Professor(a)**, em 17/07/2024, às 18:30, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Caroline da Rosa Pinheiro**, **Professor(a)**, em 19/07/2024, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-U f (www2.u f.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1790531** e o código CRC **92661BB9**.

RESUMO

O presente estudo busca argumentar acerca do surgimento de um sistema econômico baseado na financeirização da tecnologia, que, a partir da manipulação e processamento dos dados, escalou a evolução do algoritmo. Se, de um lado, o bem mais valioso da atualidade (o dado) compõe uma forma de dominação pautada na opacidade, aqui entendida como a ausência de autodeterminação informativa do titular sobre seus dados; de outro, a forma de produção deste valor (o algoritmo), maximizada pelo poder computacional mais tecnológico do cenário mundial (a inteligência artificial), se assenta sobre bases de inescrutabilidade e incompreensibilidade. Com seu poder digital, a má utilização de tais elementos passam a se agigantar como potenciais ameaças à democracia como a conhecemos. O objetivo principal deste estudo é o de compreender se, e, se sim, de que forma o Direito vem compatibilizando os novos fatores de produção aos pilares democráticos. A pesquisa se justifica pela utilização contínua e constante de tais tecnologias, a despeito das potenciais consequências dessa utilização. Os objetivos específicos estão diretamente relacionados ao objeto delimitado como tema, sendo eles: (i) arguir sobre a importância dos dados como os maiores ativos financeiros da atualidade, apresentando elementos que evidenciem a cleptocracia de dados; (ii) apresentar a ausência de autodeterminação informativa do titular dos dados e sua relação com a dominação advinda do poder sobre dados; (iii) conceituar algoritmo e analisar seu papel no extrativismo e monetização de dados no capitalismo de vigilância; (iv) conceituar a inteligência artificial como forma mais avançada da utilização do algoritmo e evidenciar, a partir de casos concretos e dados estatísticos, se os cenários já apresentados por especialistas da área tendem a se concretizar ou não na atualidade; (v) apresentar a inescrutabilidade e opacidade dos sistemas de IA como óbices à *accountability* e, em última instância, como ameaças à democracia diante da modernização dos fatores de produção; (vi) analisar fontes jurisprudenciais e legislativas do Direito a fim de compreender se há dificuldades hermenêuticas de se compatibilizar as bases principiológicas da IA e do uso de dados aos pilares democráticos. Parte-se da hipótese de que se não houvesse compatibilização hermenêutica entre os princípios basilares do Estado Democrático de Direito e a financeirização-dataficação algorítmica este último seria proibido. Havendo compatibilização, para que os incidentes continuassem a se perpetuar da forma que se vê na atualidade, parte-se da premissa de que, para além das dificuldades de aplicação prática, podem existir desajustes interpretativos das fontes do Direito. Utiliza-se uma metodologia qualitativa, valendo-se das pesquisas bibliográfica e documental, realiza-se uma revisão bibliográfica pautada numa análise de *corpus.*

Palavras-chave: inteligência artificial; dados; *accountability* algorítmica; democracia; autodeterminação informativa.

SUMÁRIO

INTRODUÇÃO

O presente estudo busca argumentar acerca do surgimento de um sistema econômico baseado na financeirização da tecnologia, que, a partir da manipulação e processamento dos dados, escalou a evolução do algoritmo - neste trabalho entendido não apenas como unidade básica de computação, mas como fator de produção do capitalismo de vigilância.

Ora, elemento indispensável ao processo produtivo da extração de dados para geração de lucro, a expressão mais avançada da tecnologia do algoritmo é a inteligência artificial.

Se, de um lado, o bem mais valioso da atualidade (o dado) compõe uma forma de dominação pautada na opacidade, aqui entendida como a ausência de autodeterminação informativa do titular sobre seus dados; de outro, a forma de produção deste valor (o algoritmo), maximizada pelo poder computacional mais tecnológico do cenário mundial (a inteligência artificial), se assenta sobre bases de inescrutabilidade e incompreensibilidade.

Quando os maiores ativos do planeta são processados por sistemas algorítmicos inescrutáveis que, sob o manto da neutralidade e racionalidade, tornam-se mentes digitais cada vez mais poderosas e imprevisíveis, estes agentes não podem ser compreendidos ou controlados sequer por seus próprios criadores. Com seu poder digital, passam a se agigantar como potenciais líderes tecnológicos não eleitos, ameaçando a democracia como a conhecemos. Esses líderes são modelos matemáticos preditivos mal concebidos e entregam sentenças algorítmicas obscuras, incontestáveis e, principalmente, irresponsabilizáveis, que terminam sendo uma caixa preta que microgerencia a sociedade moderna (O'NEIL, 2021).

No contexto democrático, no qual valores como o da transparência e o da igualdade se apresentam como alicerces para a efetivação do Estado Democrático, o objetivo principal deste estudo é o de compreender se, e, se sim, de que forma o Direito vem compatibilizando os novos fatores de produção aos pilares democráticos.

Este estudo se justifica porque, a despeito de todas as preocupações teóricas envolvendo a má utilização das IAs, no geral, e dos modelos de aprendizagem profunda, em específico, e a má coleta, gestão e tratamento dos dados que alimentam seu processo algorítmico[1]; das possíveis ameaças sociais e cenários advindos da evolução de tais modelos

[1] Citam-se, aqui, aqueles que serão apresentados por este estudo: (i) a falta de autodeterminação do titular dos dados, que raramente compreende totalmente as consequências do fornecimento de suas informações; (ii) os inúmeros ataques cibernéticos arquitetados com o fim de usurpar dados; (iii) os escândalos de vazamento de dados, especialmente aqueles associados a *big datas*; (iv) a utilização dos dados para a construção de perfis psicográficos dos usuários, o que gera, por exemplo, a possibilidade da criação de publicidades predatórias; (v) a

e dos inúmeros questionamentos e discussões que orbitam em torno de sua aplicação ideal e ética, a sociedade moderna continua, na prática, utilizando tais tecnologias e sofrendo as consequências dessa utilização.

Os objetivos específicos estão diretamente relacionados ao objeto delimitado como tema, sendo eles: (i) arguir sobre a importância dos dados como os maiores ativos financeiros da atualidade, apresentando elementos que evidenciem a cleptocracia de dados; (ii) apresentar a ausência de autodeterminação informativa do titular dos dados e sua relação com a dominação advinda do poder sobre dados; (iii) conceituar algoritmo e analisar seu papel no extrativismo e monetização de dados no capitalismo de vigilância; (iv) conceituar a inteligência artificial como forma mais avançada da utilização do algoritmo e evidenciar, a partir de casos concretos e dados estatísticos, se os cenários já apresentados por especialistas da área tendem a se concretizar ou não na atualidade; (v) apresentar a inescrutabilidade e opacidade dos sistemas de IA como óbices à *accountability* e, em última instância, como ameaças à democracia diante da modernização dos fatores de produção; (vi) analisar fontes jurisprudenciais e legislativas do Direito a fim de compreender se há dificuldades hermenêuticas de se compatibilizar as bases principiológicas da IA e do uso de dados aos pilares democráticos.

Parte-se da hipótese de que se não houvesse compatibilização hermenêutica entre os princípios basilares do Estado Democrático de Direito e a financeirização-dataficação algorítmica este último seria proibido. Havendo compatibilização, para que os incidentes continuassem a se perpetuar da forma que se vê na atualidade, parte-se da premissa de que, para além das dificuldades de aplicação prática, podem existir desajustes interpretativos das fontes do Direito.

Compreendendo os dados como os maiores ativos da modernidade, no primeiro capítulo, são apresentados os conceitos de colonialismo digital e ditadura de dados: parte-se da premissa de que os mais variados exemplos de tentativas de obtenção de poder sobre os dados representam uma nova forma de dominação na era digital. Dessa forma, o poder extrínseco advindo da ausência de autodeterminação informativa do titular dos dados e pautado na financeirização-dataficação algorítmica representa a modernização do capital. Partindo-se da compreensão de que o algoritmo representa a unidade básica de computação

disseminação de notícias falsas (*fake news*), especialmente após a sofisticação da aprendizagem de máquina e do *deepfake*; (vi) o viés algorítmico; (vii) a inescrutabilidade do modelo e a consequente impossibilidade da *accountability* algorítmica. Todos esses problemas, como também já discutido anteriormente, podem se agigantar, especialmente se considerados em conjunto, como ameaça à continuidade da democracia contemporânea como a conhecemos.

e a fonte, a partir da Revolução 4.0, do extrativismo de dados no ciberambiente, apresentam-se conceitos como os de algorítmicos de destruição em massa, *machine learning*, *deep learning* e inteligência artificial para traçar cenários de utilização da IA e questionar definições frequentemente atreladas aos algoritmos: imparcialidade, objetividade, neutralidade matemática, previsibilidade e decisionismo racional e probabilístico..

Apresenta-se o conceito de *accountability*. A inescrutabilidade e opacidade dos algoritmos mais avançados existentes na contemporaneidade levam à impossibilidade de concretização de princípios caros ao Estado Democrático de Direito e, em última instância, da responsabilização jurídica. Dessa forma, o poder digital de modelos inteligentes que processam os maiores ativos do planeta passa a ameaçar a democracia moderna.

No segundo capítulo, o conceito de extrativismo algorítmico de dados é explorado a partir de estatísticas e da explanação de casos reais que demonstram a má utilização da inteligência artificial. Se, no capítulo 1, a pretensão do estudo foi trazer cenários esboçados por especialistas da literatura sobre o tema; neste capítulo, a pretensão é de evidenciar o que verdadeiramente vem se cumprindo após os estudos realizados. Os exemplos partem do cenário global para o cenário nacional, no geral, e se estendem para a utilização da IA na seara jurídica de cada âmbito, mais especificamente. A fim de compreender de que forma os juristas vêm enfrentando o tema da utilização da IA, realiza-se uma revisão do estado da arte sobre as principais legislações existentes, nos cenários jurídicos nacional e internacional, que orbitam os temas da inteligência artificial e do extrativismo de dados.

No terceiro capítulo, apresenta-se a metodologia deste estudo. Partindo-se de uma metodologia qualitativa, valendo-se das pesquisas bibliográfica e documental, realiza-se uma revisão bibliográfica pautada numa análise de *corpus.*

No quarto capítulo, faz-se uma análise quantitativa do arcabouço legal constituído, a fim de se evidenciar o número de documentos analisados, além do número de princípios encontrados e o eventual recorte daqueles mais frequentemente citados pela literatura específica. Faz-se, ainda, uma análise qualitativa do arcabouço legal constituído, a fim de se questionar se, considerando os princípios selecionados, haveria consenso hermenêutico entre os documentos que os citam.

Por fim, no capítulo cinco e nas conclusões, apresenta-se uma reflexão acerca das principais dificuldades interpretativas levantadas, que despontam como potenciais razões para a não compatibilização, pelo Direito, entre os princípios democráticos e o arcabouço principiológico das distintas fontes jurisprudenciais. São aventados, ainda, possíveis

imbróglios jurídicos advindos de tal dissenso e que podem contribuir para a fossilização da problemática, como a inescrutabilidade da motivação de eventuais decisões judiciais humanas sobre o tema.

1 COLONIALISMO DIGITAL E A DITADURA DOS DADOS

É inegável que a contribuição das Tecnologias da Informação e Comunicação (TICs) transformou os contornos sociais da sociedade moderna. No entanto, a gleba social contemporânea não está apenas "em rede", como definiria Castells (2016), de forma que não se trata apenas da comunicação operacionalizada num novo formato de nós interconectados, mas, mais do que isso, é um novo desdobramento da sociedade do espetáculo de Debord (1997).

Para Debord (1997), o espetáculo seria não apenas um conjunto de imagens, mas uma relação social entre as pessoas, mediatizada por imagens - uma representação divisora de classes, como também defenderia Bordieu (1979), em "A Distinção: crítica social do julgamento". Tal representação teria sido engendrada pelo fetichismo da mercadoria e seria caracterizada (como forma-mercadoria) pelo quantitativo.

Dessa forma, tem-se uma sociedade do consumo marcada, de uma lado, pela economia informacional, global e em rede (CASTELLS, 2016), já que está organizada em redes de conexões globais e a partir de sua capacidade de gerar, processar e aplicar de maneira eficiente a informação; e, de outro, pela cultura de um espetáculo caracterizado pelo fetichismo da abundância da mercadoria.

Ocorre que, quando os indivíduos estão conectados por uma rede, é possível e muito provável que passem a influenciar o comportamento e as decisões uns dos outros. Isso ocorre por conta tendência humana ao pertencimento tribal (MAFFESOLI, 2004) e de forma quase involuntária, por um efeito epistêmico: porque somos mais pensados do que pensamos e porque somos mais agidos do que agimos (FOUCAULT, 2002).

É assim que as denominadas "cascatas de informações" da era informatizada dão origem a uma série de processos sociais nos quais as redes agregam comportamentos individuais, produzindo resultados coletivos (KLEINBERG; EASLEY, 2010, p. 483). Um desses processos, e talvez o mais importante e debatível, é o da superexposição.

A tecnologia pós-moderna e suas mídias sociais, abarcadas por um relacionismo galopante, funcionam a partir, principalmente, da publicização do indivíduo como forma de inserção social (AMARAL, 2016). Para Maffesoli (*apud* SAYURI, 2014), isso acontece porque, nesse novo tipo de sociedade do espetáculo, as mídias sociais são o meio e a mensagem da representação a que se referia Debord (1997).

Neste cenário, de uma sociedade inteira, em rede e se superexpondo, surge o denominado superinformacionismo, qual seja, a grande "massa de informações sobre tudo

e sobre todos, queiram ou não estar naqueles conjuntos de dados ou informações" (RULLI JUNIOR; RULLI NETO, 2013), que concretiza o fenômeno que previu Lévy, em 1999, de que a perspectiva da digitalização geral das informações tornaria o ciberespaço o principal canal de comunicação e suporte de memória da humanidade a partir do início do próximo século.

Esse sistema se retroalimenta, já que, à medida que desaparece a memória tradicional, "nós nos sentimos obrigados a acumular religiosamente vestígios, testemunhos, documentos, imagens, discursos, sinais visíveis do que foi, como se esse dossiê cada vez mais prolífero devesse se tornar prova em não se sabe que tribunal da história" (NORA, 1987, p. 15). O resultado não poderia ser diferente:

> Nenhuma época foi tão voluntariamente produtora de arquivos como a nossa, não somente pelo volume que a sociedade moderna espontaneamente produz, não somente pelos meios técnicos de reprodução e de conservação de que dispõe, mas pela superstição e pelo respeito ao vestígio (NORA, 1987, p. 15).

No entanto, para além de tornar o ciberespaço o principal canal de comunicação da era, outro fenômeno é engendrado pelo superinformacionismo: o capitalismo de vigilância (*surveillance capitalism*). Como afirma Dal Bello (2011), "a exposição generalizada da intimidade dá margem a novos modelos de exploração das informações pessoais ali depositadas (à revelia de autorização prévia)", tornando a visibilidade, a vigilância, a identidade e a indexação completamente indiscerníveis. A extração de capital a partir da gestão de dados que são, muitas vezes, fornecidos gratuitamente por usuários a empresas de tecnologia, é o que define o que se convencionou chamar de capitalismo de vigilância (ZUBOFF, 2021).

Essa extração de valor parte, muitas vezes, de uma operação denominada *data mining.* A mineração de dados, em tradução livre para o português, consiste na descoberta de padrões de relacionamentos significativos entre informações a partir da aplicação de técnicas de análise de dados (como a classificação, o agrupamento, a regressão, a análise de associação e a mineração de sequências) (HAN, PEI, TONG, 2022). Tais informações podem não aparentar utilidade, *prima facie*, no entanto, quando retiradas da perspectiva unitária, principalmente se pertencentes a uma grande base de dados - as denominadas *Big Datas* - oferecem potencial preditivo e perfilizador: o agrupamento dos dados pessoais dos usuários com o fim de extrair informações detalhadas sobre seus titulares, por exemplo, é

uma prática comum no mundo do *Business Analytics*, por ser uma técnica de identificação de padrões de consumo (BIONI, 2019).

As *big techs*, consideradas grandes empresas de tecnologias que, em regime de monopólio, coletam e tratam os dados dos usuários, são conhecidas pela utilização da análise de negócios, já que essa coleta seria capaz, em última instância, de traçar perfis de consumo, estilos de vida, preferências pessoais e políticas, influenciando um grande mercado digital (BARROS, 2021).

Esse agrupamento de informações é conhecido pela sigla PII (*personally identifiable information*), significando, em português, informações pessoais identificáveis, ou seja, que podem ser usadas para descobrir a identidade de um indivíduo[2]. Na Ciência de Dados, é o metadado que permite tal identificação - vocábulo que significa, por definição própria, um dado que fornece informações sobre outros dados. Apesar de não existir um conceito único que os defina, podem ser compreendidos, segundo a literatura específica majoritária, como "dados sobre dados" (VENTURA, DE ALBUQUERQUE SIEBRA, LIMA, 2013).

Existem diferentes tipos de metadados: os descritivos, que descrevem o conteúdo e contexto dos dados; os estruturais, que descrevem a estrutura ou o relacionamento entre os dados, ou seja, a hierarquia ou organização entre os elementos; os administrativos, que descrevem informações relacionadas à gestão dos dados, como direitos autorais, controle de acesso e política de preservação; e os de preservação, que descrevem informações como o formato de arquivo e estratégias de backup. Dessa forma, seria possível argumentar que os metadados exercem uma função importante para o gerenciamento e a eficiência e efetividade na utilização dos dados (QIN, ZENG, 2020).

Essa função, no entanto, somente pode ser operacionalizada a partir da descrição de características essenciais das informações, como formato, conteúdo, estrutura, proveniência e contexto, o que pode ser utilizado também para identificar tais dados[3] (QIN, ZENG, 2020, p. 2, tradução da autora).

Não por acaso, autores como Frazão (2019), Lira (2014) e Vieira (2007) alertam para o fato de que tais mecanismos de agrupamento podem representar um risco às liberdades individuais e à própria democracia, já que transformam a conduta num modelo

[2] Disponível em: https://www.ibm.com/br-pt/topics/pii#:~:text=As%20informa%C3%A7%C3%B5es%20pessoalmente%20identific%C3%A1veis%20(PII,ou%20endere%C3%A7o%20de%20e%2Dmail. Acesso em 10 de dez. de 2023.

[3] "Metadata is data that provide information about other data. It describes essential characteristics of data such as format, content, structure, provenance, and context. Metadata can be used to identify, locate, manage, and preserve data as well as to enhance the access, usability, and interoperability of data".

lucrativo de negócio. A possibilidade do cruzamento de dados pessoais, muitas vezes oferecidos pelos próprios usuários, poderia, dessa forma, agigantar-se como uma ameaça ao direito à privacidade.

É o caso do emblemático episódio ocorrido com a Target, em meados de 2012. A empresa mantinha uma base de dados com padrões de consumo que eram capazes de traçar modelos preditivos. Um exemplo é o evento de uma gravidez, que estava estatisticamente associado à compra de loções sem essência e de suplementos alimentares. A empresa criou uma lista com 25 produtos que uma mulher grávida costuma comprar e era possível estimar, com essas informações, uma probabilidade de 0 a 100% e o estágio da gravidez (em semanas). Essa prática servia para que a loja pudesse enviar cupons de descontos e ofertas personalizadas para as potenciais mães. No entanto, em uma das redes varejistas, em Minneapolis, uma adolescente recebeu um cupom de desconto relacionado à gravidez antes mesmo de saber que estava grávida, o que lhe gerou sérios desconfortos familiares (O GUIA FINANCEIRO, 2019).

Casos como estes, em que cadastros e inscrições "gratuitas" dão acesso a determinado benefício - como é o caso de informar algum dado pessoal para ter descontos em farmácias ou acesso ao Wi-Fi de lugares públicos - deram ensejo à criação da expressão "Se você não está pagando pelo produto, você é o produto" (O DILEMA DAS REDES, recurso online), disseminando o entendimento de que o próprio dado é o valor do bem.

O grande imbróglio existente nessa seara é o de que raramente o consumidor é informado da utilização ou da forma de tratamento que será dada àquela informação, motivo pelo qual a faculdade de o sujeito construir sua própria esfera particular, exercendo o controle sobre suas próprias informações, torna-se objeto sujeito à tutela do direito contemporâneo à privacidade (RODOTÁ, 2018).

Para além do problema da falta de autodeterminação informativa, os titulares do que se tornaram os ativos mais caros do mundo - os dados - passaram a ser bombardeados por infinitivos tipos de ataques arquitetados para roubar a mercadoria mais valiosa do novo modelo econômico. E as consequências são catastróficas: *ransomware, phishing, backdoor, cryptojacking, spoofing, man-in-the-middle (MitM), denial of service (DoS), direct memory access (DMA), decoy, zeroday*[4] e outras tantas técnicas de cibercrimes trouxeram consigo

[4] O *ransomware* é um tipo de prática que se utiliza de um *malware* - um software malicioso - para manter dados reféns até que seu titular concorde em pagar um resgate ao invasor; é também conhecido como sequestro de dados. O *phishing* é um tipo de estelionato: o criminoso dá golpes, através de engenharia social, para obter dados do titular. *Backdoor* significa, em tradução livre, "porta dos fundos" e é uma forma de contornar a segurança do sistema para ter acesso ao controle geral de dados e informações. O *cryptojacking* é um ataque a dispositivos com o fim de fazer mineração de criptomoedas. O *spoofing* é um ataque cibernético que consiste na

as maiores polêmicas de vazamento e roubo de dados para a construção de perfis psicográficos, além de diversos escândalos envolvendo a disseminação de *fake news* e a falta de transparência e *accountability* no que se refere ao tratamento de dados e metadados.

Levantamentos de empresas que se dedicam ao estudo da cibercultura e do gênero criminal inerente à esta seara, o cibercrime, apontam para um aumento drástico dessa modalidade criminal no cenário nacional durante os últimos anos (G1, 2021). A Central Nacional de Denúncias de Crimes Cibernéticos, uma parceria da ONG "Safernet" Brasil com o Ministério Público Federal (MPF), divulgou dados que demonstram que o total de cibercrimes cometidos no ano de 2020 (156.692) foi o maior da série histórica desde que o levantamento começou, em 2014, sendo, ainda, o dobro do total do ano de 2019.

No ano de 2021, os números não foram muito animadores: nos primeiros oito meses do ano, o Brasil teve um aumento de 23% nos casos de cibercrimes em comparação ao mesmo período do ano anterior (FORUM KONFERENCIA @CASA KASPERSKY, 2021). Além de liderar os rankings de *ransomware* e *phishing,* o país registrou, nesse período, 481 milhões de tentativas de infecção, o que equivale a 1.395 tentativas por minuto (OLHAR DIGITAL, 2021), e os órgãos públicos, como os cidadãos, se apresentaram igualmente frágeis à tendência dos ataques.

Na esteira do aumento exponencial de cibercrimes, estão os vazamentos de dados pessoais sensíveis, aliados ou não à disseminação de notícias falsas (*fakes news)*: em 2013, o vazamento de dados, inclusive bancários, de mais de 152 milhões de nomes e senhas, além de 2,8 milhões de números de cartões de crédito de usuários da empresa Adobe, além do vazamento de pelo menos 40 milhões de dados de cartões e 70 milhões de outros dados dos clientes da franquia Target; em 2014, o roubo de cerca de 56 milhões de números de cartões de créditos da Home Depot; em 2016, o vazamento dos dados de mais de 57 milhões de usuários da Uber (destes, 196 mil brasileiros); em 2017, o vazamento de mais de 1,4 bilhão de nomes de usuários e senhas de sites como Netflix, LinkedIn, MySpace, Las.Fm e YouPorn, além da disseminação do *ransomware* WannaCry, que infectou mais de

falsificação de endereços para que o invasor possa se passar pelo usuário e ter acesso às suas informações. O *Man-in-the-middle* (em tradução livre, "homem no meio") ocorre quando um hacker - ou invasor - se insere entre os meios de comunicação de um usuário para interceptar seus dados. O *DoS* é um ataque que faz com que determinado sistema fique indisponível enquanto os invasores têm acesso aos dados daquele diretório. O DMA é um ataque de acesso direto à memória do sistema. O *decoy* é o ataque realizado através de um programa legítimo hackeado, fazendo com que o usuário entre no "programa espelho" com seus dados reais. O *zeroday* (em tradução livre, "dia zero") é um ataque a partir das falhas de seguranças (ou *bugs*) de aplicações recém-lançadas (FORTINET, 2022).

230 mil sistemas; em 2021, o vazamento de 8.4 bilhões de senhas de usuários da rede social RockYou (ARAÚJO, 2020; RODRIGUES, 2021).

Um dos vazamentos de dados pessoais que ganhou mais repercussão foi o do Facebook, em 2019, no qual mais de 87 milhões de usuários (destes, mais de 443 mil brasileiros) tiveram seus dados pessoais publicados gratuitamente em fóruns hacker. Informações incluindo nome, endereço, número de telefone, data de nascimento e e-mail foram obtidas a partir de uma ferramenta existente, à época, na rede social e que permitia importar os dados dos usuários e associá-los, criando metadados. O caso do vazamento dessa plataforma é tido como um dos mais chocantes também pela acusação de conexão do Facebook com a eleição de Donald Trump, em 2016 - quando a empresa Cambridge Analytica coletou e tratou dados do Facebook com o fim de manipular eleitores e influenciar os resultados eleitorais norte-americanos (Senha Segura, 2021).

Não por acaso, diante de tantas controvérsias, a literatura especializada passou a utilizar os termos colonialismo digital e ditadura de dados para se referir à nova forma de dominação na era da tecnologia: o poder sobre os dados (SIQUEIRA, 2019).

Essa nova forma de imperialismo está intrinsecamente relacionada ao poder de dominação advindo da falta de autodeterminação informativa no âmbito do maior valor econômico da contemporaneidade: os dados. O conceito de financeirização-dataficação algorítmica (DARDOT; LAVAL, 2016) parte do entendimento de que o extrativismo de dados pessoais na era moderna não configura mera coleta de informação, senão extração de recursos (SRNICEK, 2016), já que o processo de documentação, filtragem, extração e acúmulo dos dados caracteriza o novo padrão sistêmico de geração de riqueza e, consequentemente, da modernização do capital (MURDOCK, 2018).

Todo esse processamento somente é possível a partir da unidade básica da computação: o algoritmo (VAN DIJCK; POELL; DE WALL, 2018).

1.1 ALGORITMOS DE DESTRUIÇÃO EM MASSA

Alguns autores, diante dessa nova era tecnológica, passaram a cunhar o termo “Quarta Revolução Industrial” ou “Revolução 4.0”. O termo se refere, segundo Klaus Schwab (2019), à "combinação de fatores estruturais (excesso de endividamento e envelhecimento das sociedades) e sistêmicos (a introdução da plataforma e das economias

sob demanda, a crescente relevância da diminuição dos custos marginais etc.)" que reescreveu os livros de economia.

Essa revolução industrial, como todas as outras, deveria apontar para uma melhoria da eficiência e da produtividade dos processos, no entanto, sua característica principal é demarcada pela presença de conceitos tecnológicos que têm potencial para um crescimento econômico nunca antes visto (SCHWAB, 2019). No entanto, como apresentado, esse não é o único potencial da alteração produzida pela Quarta Revolução.

Dentre as novas definições do ciberambiente (ou também denominado ciberespaço), ou seja, pelo "novo espaço de comunicação proporcionado pela interconexão mundial de computadores e das memórias dos computadores" (LEVY, 2010), Cathy O'Neil, matemática, cientista de dados e escritora, chama atenção, no livro Algoritmos de Destruição em Massa (2021), para seu potencial destrutivo.

A autora afirma, em sua obra, abordando também sua experiência profissional e pessoal, que o processamento contínuo de dados obtidos de redes sociais ou sites de e-commerce vêm enfocando cada vez mais os seres humanos:

> Matemáticos e estatísticos estavam estudando os nossos desejos, movimentações e poder de compra. Eles previam nossa credibilidade e calculavam nosso potencial enquanto estudantes, trabalhadores, amantes e criminosos. Esta era a economia do Big Data, os imensos volumes de dados, e ela prometia ganhos espetaculares. Um programa de computador poderia vasculhar milhares de currículos ou pedidos de empréstimo em um segundo ou dois e ordená-los em listas impecáveis, com os candidatos mais promissores no topo. Isso não apenas economizava tempo, mas também era vendido como algo imparcial e objetivo. Afinal, não envolvia humanos preconceituosos cavoucando resmas de papel, apenas máquinas processando números frios. Por volta de 2010, a matemática impunha-se como nunca nas questões humanas, e o público amplamente a saudava (O'NEIL, 2021, p. 6).

A alcunha "algoritmos de destruição em massa" ou "armas de destruição matemática" (ADM) foi criada pela autora para caracterizar situações em que, sob o manto da imparcialidade, objetividade e neutralidade matemática, algoritmos de policiamento preditivo acentuam desigualdades e injustiças.

Os algoritmos podem ser compreendidos, dentro da lógica computacional, como conjuntos de passos finitos e organizados que, quando executados, resolvem um determinado problema (MANZANO, 2000). Para resolver determinado problema, é comum que a ciência de dados utilize uma representação abstrata de um sistema real ou conceitual, a fim de compreender o comportamento desse sistema e prever seu desempenho (O'NEIL,

2021). Essa representação abstrata, normalmente pautada em dados de um histórico, é denominada modelo.

A aprendizagem de máquina, ou *machine learning*, é o subcampo da IA que se concentra em desenvolver algoritmos e modelos que aprendam a partir de dados, ou seja,

> O *machine learning* envolve o uso de algoritmos que permitem que um sistema faça previsões ou decisões com base em dados, sem essas previsões ou decisões que precisam ser explicitamente programadas. Em vez disso, os algoritmos são projetados para encontrar padrões nos dados e ajustar seus parâmetros para melhorar sua capacidade de fazer previsões ou decisões (GÉRON, 2022).

Os algoritmos, portanto, treinam um modelo de aprendizagem de máquina com base em um conjunto de dados para que esse modelo possa ser usado para fazer previsões ou tomar decisões sobre novos dados. Estes são modelos preditivos porque analisam dados de um histórico e calculam o posicionamento com maior probabilidade de sucesso (O'NEIL, 2021)

Os algoritmos mais comuns para o treinamento da aprendizagem de máquina são o aprendizado por reforço, o não supervisionado, o semi-supervisionado e o supervisionado (BRAZDIL, GAMA, 2010).

No algoritmo de aprendizado por reforço, o objetivo é maximizar as recompensas totais ao longo do tempo e o modelo é treinado para tomar ações a partir de um feedback em tempo real de recompensa ou penalidade. A aprendizagem por reforço tem aplicações em jogos de computador, como é o caso da IA *AlphaGo*, desenvolvida pela *Google DeepMind*, que usa esse modelo de aprendizagem para jogar o jogo de tabuleiro *Go* (BRAZDIL, GAMA, 2010).

Já no algoritmo de aprendizagem não supervisionado, não há dados rotulados e o modelo é treinado a partir do agrupamento dos dados em *clusters*, identificando relações entre estes. A aprendizagem não supervisionada possui aplicações em reconhecimento de imagens, classificação de documentos e recomendação de conteúdo em redes sociais (BRAZDIL, GAMA, 2010).

Por fim, no algoritmo de aprendizagem supervisionada, os dados de entrada são rotulados com uma resposta correta e o modelo é treinado a partir de um conjunto de dados de treinamento rotulados para prever a resposta para novos dados de entrada ainda não rotulados. Apesar de poder ser utilizado em outras categorias de aprendizagem de máquina, uma das técnicas mais conhecidas do aprendizado supervisionado é o *deep learning*, um

“conjunto de técnicas de aprendizado de máquina baseadas em redes neurais profundas, que são capazes de aprender representações complexas dos dados de entrada”. Essas técnicas se diferenciam das redes neurais tradicionais porque apresentam arquiteturas mais profundas e com camadas especializadas na extração de características relevantes dos dados; normalmente são utilizadas para tarefas como reconhecimento de voz, visão computacional, processamento de linguagem natural, entre outras (SILVA, SPATTI, 2017).

Existe, ainda, um modelo híbrido e mais moderno de aprendizagem semi-supervisionada que trabalha com poucos exemplos rotulados na entrada e que não será examinado profundamente neste trabalho, por conta do recorte do estudo (MATSUBARA, 2004).

Na atualidade, estes modelos parecem ser os que mais se aproximam do conceito de inteligência artificial, considerada a expressão mais avançada e tecnológica da unidade básica de computação na modernidade.

Uma das referências mais prestigiadas na ciência de dados acerca do conceito de inteligência artificial, o livro *Artificial Intelligence: A Modern Approach*, de Stuart Russell e Peter Norvig, publicado em 1995 e atualizado em 2021, descreve a IA como "o ramo da ciência da computação que se preocupa com a criação de máquinas que podem executar tarefas que exigem inteligência humana" (RUSSELL, 2021), como o reconhecimento de fala, a compreensão da linguagem natural, o aprendizado e a tomada de decisão.

Mais recentemente, surgiu um sub ramo da IA denominado Inteligência Computacional, cujo objeto são os mecanismos adaptativos que permitem um comportamento inteligente da máquina em ambientes complexos: tais mecanismos são pautados em paradigmas da capacidade humana, como o de aprender, generalizar, abstrair, descobrir e associar (SALOMÃO, 2021).

Não por acaso, o teste de Turing, concebido por Alan Turing em 1950, avaliava que uma máquina poderia ser considerada inteligente se pudesse fingir ser humana com sucesso para um observador experiente. O teste funcionava através de um teletipo (para evitar exigir que a máquina imitasse a aparência ou a voz do humano envolvido) e o observador deveria interagir com a máquina e com o humano, a fim de distingui-los (TURING, 2009).

O *deep learning* é concebido a partir de unidades básicas de processamento capazes de imitar o comportamento do cérebro humano, como o são os neurônios artificiais - capazes de receber informações de outras unidades, processá-las e transmitir um resultado para outras unidades; e as redes neurais artificiais, sistemas de unidades interconectadas capazes de aprender e se adaptar a partir de determinado *input* (dados de entrada) para

reconhecer padrões ou resolver problemas de classificação (MCCULLOCH, PITTS, 1943). Por isso, podem ser considerados os modelos que, na contemporaneidade, representam melhor a máquina inteligente de Turing.

No mundo moderno, o que era entendido como um agente inteligente analítico vem se tornando, cada vez mais, um agente inteligente generativo: ou seja, se antes a IA era utilizada para tarefas de análise preditiva e reconhecimento de imagem e fala, hoje é utilizada para a criação de novos elementos. Num momento anterior, a humanidade conheceu a utilização da IA para tarefas específicas, como para o processamento de linguagem natural (PLN), ação voltada para a compreensão, interpretação e manipulação da linguagem humana, ou para a mineração de textos, ação voltada para a busca de termos relevantes ou de padrões em grandes volumes de dados; na modernidade, a IA é capaz de criar vídeos, áudios, imagens e quaisquer outras mídias digitais (GUIDI; TIMPONE, 2023).

Esses avanços tecnológicos permitiram a criação, pelas IAs, de conteúdos hiper-realistas e que deixam poucos ou quase nenhum vestígio de manipulação (CHAWLA, 2019): são os denominados *deepfakes*, mídias que, ao mesclar, combinar, substituir e sobrepor outras mídias, criam conteúdos que parecem autênticos (ALEXANDROU; MARA, 2018). E, apesar de espalhar esse tipo de mídia falsa (criando as *fake news*) seja fácil, corrigir esse registro é extremamente difícil (DE KEERSMAECKER; ROETS, 2017).

Assim, quando a mimetização da atividade nervosa, não mais como um processo metabólico, mas sim como um circuito de neurônios interagindo como uma rede lógica de interruptores, passa a ser tratada pelos métodos da lógica simbólica e a atividade nervosa envolvida nos processos de percepção, memória e raciocínio passa a ser descrita por um cálculo algébrico[5], surge, para além de um questionamento filosófico acerca da natureza da consciência e da distinção entre mente e cérebro, uma preocupação com o futuro da humanidade.

Isso porque, diante da escalada dos modelos de aprendizagem de máquina, não basta a aplicação de um teste unilateral para a aferição de inteligência. Ou seja, um teste como o de Turing deve ser considerado também parcial: uma máquina que passa no teste deve ser considerada inteligente, mas uma máquina ainda poderia ser considerada inteligente sem saber o suficiente sobre humanos para imitá-los. Há que se considerar um debate acerca das

[5] "We shall show that all nervous activity, when it is not mere metabolism, is resolvable into circuits of interacting neurons which are themselves logical networks of switches, exciting and inhibiting each other. These circuits can be treated by the methods of symbolic logic, and therefore the nervous activity involved in the processes of perception, memory and reasoning can be described by an algebraic calculus" (MCCULLOCH, PITTS, 1943, s/p).

restrições do conhecimento do observador sobre a IA, porque “algumas pessoas são facilmente conduzidas a acreditar que um programa bastante estúpido é inteligente” (MCCARTHY, 2007, tradução da autora).

Warren McCulloch e Walter Pitts, ainda em 1943, já previam o nascimento desse e de outros possíveis embates envolvendo a criação de um modelo de inteligência capaz de imitar o comportamento da inteligência humana e, a despeito das posteriores sucessivas criações cinematográficas distópicas envolvendo a IA, tais questionamentos seguem orbitando a temática sem um consenso definitivo no mundo contemporâneo. Uma das principais preocupações com o rápido desenvolvimento dos modelos inteligentes diz respeito aos riscos existenciais da IA para a humanidade.

Nessa toada, ainda em 2008, o Painel Presidencial AAAI 2008-09 sobre Futuros de IA de Longo Prazo, grupo de estudo organizado pela Associação para o Avanço da Inteligência Artificial (AAAI) com o objetivo de examinar possíveis cenários futuros para a inteligência artificial (IA) em um horizonte de 20 a 100 anos, confeccionou o "Relatório do Painel Presidencial AAAI 2008-09 sobre Futuros de IA de Longo Prazo", publicado em 2009. O relatório dividiu os possíveis cenários em cinco hipóteses.

Num cenário otimista, a IA leva a avanços significativos, possibilitando uma melhor compreensão do cérebro humano e da cognição; num cenário pessimista, a IA se torna uma ameaça para a humanidade, levando a uma catástrofe global ou a uma perda significativa de empregos e desigualdades socioeconômicas; num cenário neutro, a IA se desenvolve de forma incremental e é amplamente utilizada em tarefas específicas; num cenário de fusão homem-máquina, a IA é integrada ao corpo humano, permitindo avanços significativos nas capacidades humanas; num cenário de descentralização, por fim, a IA é distribuída em sistemas em rede e colaborativos, com ênfase na autonomia e na capacidade de adaptação em tempo real[6].

Alguns anos depois, em 2014, é fundado o *Future of Life* (FLI), uma organização sem fins lucrativos que se dedica a promover segurança no desenvolvimento de tecnologias avançadas e, em 2015, o instituto publica uma carta aberta sobre a criação de sistemas robustos de IA[7]. A *Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter* (2015), “Prioridades de pesquisa para inteligência artificial robusta e benéfica: uma carta aberta”, em tradução livre para o português, visa o enfoque em pesquisas de IA

[6] Disponível em https://www.aaai.org/Organization/Organization/Future_of_AI_Panel_Report.pdf. Acesso em 02 de maio de 2023.
[7] Disponível em: https://futureoflife.org/fli-open-letters/. Acesso em 02 de maio de 2023.

que criassem sistemas não apenas robustos, mas, principalmente, benéficos para a humanidade.

O documento parece se posicionar de forma otimista em relação à utilização de agentes inteligentes, narrando benefícios da adoção de representações probabilísticas e teóricas de decisão e métodos de aprendizado estatístico, que levaram a uma integração e fertilização cruzada entre IA, aprendizado de máquina, estatística, teoria de controle, neurociência e outros campos. Um de seus excertos menciona:

> O progresso na pesquisa de IA torna oportuno focar a pesquisa não apenas em tornar a IA mais capaz, mas também em maximizar o benefício social da IA. Tais considerações motivaram o Painel Presidencial AAAI 2008-09 sobre Futuros de IA de Longo Prazo e outros projetos sobre os impactos da IA, e constituem uma expansão significativa do campo da própria IA, que até agora tem se concentrado amplamente em técnicas que são neutras em relação à propósito (FLI, 2015, recurso online, tradução da autora).

A carta conceitua os agentes inteligentes como sistemas que percebem e agem em algum ambiente e relaciona a definição de inteligência, nesse contexto, a noções estatísticas e econômicas da racionalidade, ou seja, à habilidade de tomar boas decisões e fazer bons planos ou inferências (FLI, 2015). Dessa forma, a inteligência do modelo estaria assentada em noções matemáticas capazes de gerar predições e, mais que isso, por se basear em técnicas neutras, parece haver uma crença de que esse tipo de inteligência não se sustentaria em juízos de valor.

Em 2017, a preocupação com um desenvolvimento seguro dos modelos inteligentes ainda é uma pauta latente e, em uma conferência realizada na *Asilomar Conference Grounds*, na Califórnia, especialistas em inteligência artificial se reúnem para propor um conjunto de 23 diretrizes para um uso mais responsável da IA. O documento, que fica conhecido como Princípios de IA de Asilomar, não é juridicamente vinculante, mas se torna um marco ético na seara de ciência de dados[8].

[8] Questões sobre a pesquisa
1) Objetivo da Pesquisa: O objetivo da pesquisa em IA não deve ser criar inteligência não direcionada, mas inteligência benéfica.
2) Financiamento da Pesquisa: Investimentos em IA devem ser acompanhados de financiamentos em pesquisa para assegurar seu uso benéfico, incluindo pesquisas espinhosas em Ciência da Computação, Economia, Leis, Ética, e Estudos Sociais, tais como:
3) Link entre Ciência e Política: Deve haver um intercâmbio construtivo e saudável entre pesquisadores de IA e formuladores de políticas.
4) Cultura de pesquisa: Uma cultura de cooperação, confiança e transparência deve ser fomentada entre pesquisadores e desenvolvedores de IA.
5) Prevenção de Corrida: Equipes que desenvolvem sistemas de inteligência artificial devem cooperar ativamente para evitar cortes nas normas de segurança.
Éticas e Valores

Dentre os 23 princípios, destacam-se, para os fins deste trabalho, a Transparência de Falha (7), a Transparência Judicial (8), a Responsabilidade (9), a Privacidade Pessoal (12) e a Liberdade e Privacidade (13), que parecem apontar para uma preocupação com a revisão e responsabilização das decisões tomadas pela IA e com a moralidade no uso dos dados apreendidos por esses modelos.

Em 24 de fevereiro de 2023, a OpenAI, empresa de pesquisa em IA fundada em 2015, por um grupo de empresários e cientistas, dentre os quais vale mencionar o célebre Elon Musk, publicou uma declaração afirmando que, a curto prazo, pretendia criar uma transição gradual entre a IA e a *Artificial General Intelligence* (AGI).

A AGI ou, em tradução livre para o português, inteligência artificial geral, seria um modelo capaz de realizar qualquer tarefa que um ser humano pudesse executar, aprendendo

---

6) Segurança: Os sistemas com IA devem ser seguros e protegidos durante toda a sua vida útil operacional, e verificáveis, quando aplicável e viável.
7) Transparência de falha: Se um sistema com IA causar dano, deve ser possível determinar o motivo.
8) Transparência Judicial: Qualquer envolvimento de um sistema autônomo na tomada de decisões judiciais deve fornecer uma explicação satisfatória passível de auditoria por uma autoridade humana competente.
9) Responsabilidade: Designers e construtores de sistemas avançados com IA são partes interessadas nas implicações morais de seu uso, abuso e ações, com responsabilidade e oportunidade de moldar essas implicações.
10) Alinhamento de valor: Sistemas com IA altamente autônomos devem ser projetados de modo que seja assegurado que seus objetivos e comportamentos serão alinhados com os valores humanos durante toda a operação.
11) Valores Humanos: Os sistemas com IA devem ser projetados e operados de modo a serem compatíveis com os ideais da dignidade humana, direitos, liberdades e diversidade cultural.
12) Privacidade Pessoal: As pessoas devem ter o direito de acessar, gerenciar e controlar os dados que geram, dado o poder dos sistemas com AI de analisar e utilizar esses dados.
13) Liberdade e Privacidade: A aplicação de IA aos dados pessoais não deve restringir de forma injustificável a liberdade real ou percebida das pessoas.
14) Benefício compartilhado: Tecnologias com IA devem beneficiar e capacitar o maior número de pessoas possível.
15) Prosperidade compartilhada: A prosperidade econômica criada pela IA deve ser compartilhada amplamente, para beneficiar toda a humanidade.
16) Controle Humano: Os seres humanos devem escolher como e se devem delegar decisões aos sistemas com IA, para realizar os objetivos escolhidos pelo homem.
17) Não-subversão: O poder conferido pelo controle de sistemas com IA altamente avançada deve respeitar e melhorar, ao invés de subverter, os processos sociais e cívicos dos quais depende a saúde da sociedade.
18) Corrida Armada com IA: Deve ser evitada uma corrida armamentista com armas autônomas letais.
Questões de Longo Prazo
19) Atenção na Capacidade: Não havendo consenso, devemos evitar fortes suposições sobre os limites superiores em futuras capacidades de IA.
20) Importância: IA avançada poderia representar uma mudança profunda na história da vida na Terra, e deveria ser planejada e administrada com cuidado e recursos proporcionais.
21) Riscos: Os riscos colocados pelos sistemas com IA, especialmente os riscos catastróficos ou existenciais, devem estar sujeitos a esforços de planejamento e mitigação proporcionais ao impacto esperado.
22) Auto Aprimoramento Recursivo: Sistemas com IA projetados para melhorar ou autorreplicar-se recursivamente de uma maneira que poderia levar a um aumento rápido da qualidade ou quantidade, devem estar sujeitos a rígidas medidas de segurança e controle.
23) Bem comum: A superinteligência só deve ser desenvolvida a serviço de ideais éticos amplamente compartilhados, e para o benefício de toda a humanidade e não de um estado ou organização (FLI, 2017, recurso online).

a partir de informações complexas e sendo capaz, ainda, de raciocinar em níveis avançados, resolver problemas, perceber e manipular objetos no mundo físico e se comunicar com outros agentes inteligentes (MCCARTHY, 1987). Vale lembrar que a aspiração a uma forma de inteligência não especializada, ou seja, que não fosse criada para executar tarefas específicas, mas que fosse capaz de desempenhar qualquer tarefa intelectual que um ser humano pudesse realizar (MCCARTHY, 2007) já era conhecida desde os anos 60[9], apesar de nunca ter sido implementada.

A OpenAI, conhecida pelo desenvolvimento de modelos de linguagem natural avançados, como o ChatGPT - que é uma versão compacta do GPT-3, modelo treinado em uma base de dados maior, afirmou, nesta mesma declaração, publicada no site da empresa, que o futuro da humanidade deveria ser determinado pela humanidade, tendo apontado, para tanto, três questões-chaves para uma conversa global sobre o tema: como governar esses sistemas, como distribuir de forma justa os benefícios que eles geram e como compartilhar o acesso de forma justa[10].

Um mês depois da declaração, o instituto FLI publica mais uma carta aberta, denominada a *Pause Giant AI Experiments: An Open Letter (2023),* "Pausa nos experimentos gigantes de IA: uma carta aberta", também em tradução livre, propondo uma pausa temporária nos experimentos gigantes em inteligência artificial para permitir que a comunidade global discutisse e estabelecesse diretrizes sobre o desenvolvimento seguro e responsável da IA (FLI, 2023):

> [...] a IA avançada pode representar uma mudança profunda na história da vida na Terra e deve ser planejada e gerenciada com cuidados e recursos proporcionais. Infelizmente, esse nível de planejamento e gerenciamento não está acontecendo, embora os últimos meses tenham visto laboratórios de IA travados em uma corrida descontrolada para desenvolver e implantar mentes digitais cada vez mais poderosas que ninguém – nem mesmo seus criadores – pode entender, prever ou controlar de forma confiável (FLI, 2023, recurso online, tradução da autora).

O documento traz à tona, ainda, questionamentos sobre a inundação de canais de informação por propagandas e informações falsas, a automatização de todos os trabalhos

[9] John McCarthy, conhecido por ter cunhado o termo inteligência artificial, ainda em 1956, e por ter trabalhado, em 1960, no primeiro programa de xadrez capaz de jogar contra humanos (o *Mac Hack VI*), foi um dos primeiros cientistas a defender a ideia (MCCARTHY, 1955).
[10] Disponível em: https://openai.com/blog/planning-for-agi-and-beyond. Acesso em 02 de maio de 2023.

humanos, a criação de uma superinteligência[11] e a perda do controle da civilização, sustentando que algumas decisões "não devem ser delegadas a líderes tecnológicos não eleitos". Por fim, assevera que sistemas poderosos de IA devem ser desenvolvidos somente quando houver confiança nos impactos positivos e pede uma pausa mínima (pública e confiável) de 6 (seis) meses no treinamento de sistemas mais poderosos que o GPT-4 (modelo de linguagem natural mais avançado da empresa OpenAI). A carta aberta foi assinada inclusive por líderes do setor tecnológico, como Stephen Hawking, Bill Gates e pelo próprio Elon Musk, proprietário da OpenAI.

Apesar da preocupação global com o potencial desastroso do mau uso da IA estar cercada de possíveis cenários em que a tecnologia se torna uma ameaça social, há, mais especificamente, um panorama jurídico alarmante: a *accountability* algorítmica.

[11] A ideia da superinteligência refere-se à suposição de que a IA poderia ultrapassar, num futuro hipotético, o nível de inteligência humano - seria a chamada singularidade. Os pilares dessa teoria estão enraizados na unidimensionalidade da inteligência e na capacidade de autoaperfeiçoamento da IA (SCHNEIDER, 2016).

2 A ACCOUNTABILITY ALGORÍTMICA E A DEMOCRACIA MODERNA

Ainda segundo O'Neil (2021), com o fenômeno da Internet,

> [...] pessoas de todo o mundo têm produzido quadrilhões de palavras sobre nossas vidas e trabalhos, amizades e forma como compramos. Ao fazê-lo, construímos de modo involuntário a mais vasta coletânea de treinamento para máquinas de linguagem natural. Conforme trocamos papel por e-mail e redes sociais, as máquinas podiam estudar nossas palavras, compará-las com outras, e deduzir algo sobre seu contexto. O progresso tem sido rápido e dramático. Esses avanços em linguagem natural abriram um filão de possibilidades para os anunciantes. Os programas "sabem" o que uma palavra significa, ao menos o bastante para associá-la com certos comportamentos e resultados, ao menos parte das vezes. Impulsionados em parte por esse crescente domínio da língua, os anunciantes podem sondar padrões mais profundos. Um programa de publicidade pode começar com os detalhes mais comuns de demografia e geografia. Mas no curso de semanas e meses ele começa a aprender os padrões das pessoas em que está mirando e fazer previsões acerca de seus próximos passos. Passa a conhecê-las. E se for um programa predatório, afere suas fraquezas e vulnerabilidades e persegue o caminho mais eficiente para explorá-las (O'NEIL, 2021, p. 74).

O monstruoso poder auferido pelas *Big Techs* passa a ser sensível na democracia quando há comprovações científicas de que postagens influenciam o comportamento de voto das pessoas: apesar de geralmente serem utilizados para os lucros de tais empresas, os dados passam a ser utilizados na política, já que cabe às políticas governamentais regulamentar os titãs da tecnologia (O'Neil, 2021).

Neste ponto, o *machine learning,* que "faz com que a máquina aprenda certas funções a ponto de conseguir agir sem a interferência humana", quando evolui para o *deep learning* e para a criação de redes neurais artificiais, se torna um poderoso artifício, já que é tido como uma tecnologia pautada numa racionalidade formal e probabilística, capaz de realizar escolhas mais eficientes, objetivas e imparciais do que as humanas, sujeitas a falhas de enviesamento (NUNES, 2018).

Ocorre, no entanto, que tal alegada neutralidade vem se demonstrando apenas aparente, já que a literatura específica diagnosticou que o algoritmo herda todo o conteúdo de seu humano-criador, inclusive o preconceito (FRAZÃO, 2018), como apontam os célebres exemplos de enviesamento algorítmico.

No caso COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), por exemplo, o algoritmo de Perfil de Gerenciamento Corretivo de Infratores para Sanções Alternativas, elaborado pela empresa Northpointe (atual Equivant) e utilizado

nos EUA para auxiliar juízes em suas decisões, apresentou enviesamento de raça. O modelo avaliava a probabilidade de reincidência dos réus e, após investigação jornalística, foi verificado que as pessoas negras que apresentavam alto risco de reincidência não eram acusadas novamente; já as pessoas brancas de baixo risco reincidiam em crimes (VIEIRA, 2019).

Outro episódio célebre envolvendo vieses algorítmicos diz respeito aos anúncios de emprego da empresa Google: um estudo empírico sobre carreiras em STEM (Ciência, Tecnologia, Engenharia e Matemática), pautado na criação de anúncios fictícios de posições de emprego e na atribuição aleatória a perfis de usuários masculinos ou femininos, identificou que quando os anúncios foram direcionados a usuários com base em informações demográficas, houve uma tendência de exibir os anúncios de carreira em STEM para um público predominantemente masculino. Além disso, os custos de anúncios voltados para mulheres eram maiores, por conta de uma menor taxa de cliques (LAMBRECHT; TUCKER, 2019).

Para além das críticas à pretensa neutralidade algorítmica, há também os questionamentos quanto aos processos deste tipo de tecnologia. Uma das questões mais controversas dos modelos de redes neurais artificiais, nessa toada, é a inócua tentativa de regular suas operações algorítmicas dirigidas por dados: diferentemente de outras tecnologias, o modelo de aprendizado profundo demonstra uma impossibilidade da explicação dos procedimentos realizados por seu sistema algorítmico, que age segundo a estrutura de dados que obtém. Essa impossibilidade, denominada na Ciência de Dados inescrutabilidade do sistema, significa que as operações realizadas nas camadas das redes neurais não podem ser compreendidas (SILVEIRA, 2020).

Em outros tipos de aprendizagem de máquina, como é o caso dos algoritmos de árvores de decisão, regressão linear ou regressão logística, isso não acontece, já que as regras para a escolha de determinado resultado são evidentes: é possível descobrir matematicamente quais nós de uma rede neural profunda foram ativados, mas não as operações realizadas pelos neurônios que modelaram o resultado (TIWARI; TIWARI; TIWARI, 2018, p. 8).

Já existem propostas para a criação de um modelo dotado de inteligência artificial explicável - o que ficou conhecido como XAI, ou seja, uma tecnologia cujas predições pudessem ser qualitativamente interpretáveis ou explicáveis pelo próprio agente inteligente (RIBEIRO *et al*., 2018). Para além dessa proposta, alguns cientistas da computação afirmam

que todo algoritmo pode ser compreendido a partir da estrutura de dados que o acompanha (SEAVER, 2019).

O antropólogo Nick Seaver (2019), no entanto, afirma que a transparência não é a solução: modelos de aprendizagem profunda são tão complexos que se tornam uma caixa-preta inviolável, tanto por conta do volume de dados que os alimenta quanto em razão do número de pessoas envolvidas em seu desenvolvimento. Tais fatores engendrariam um nível de enredamento que seria um óbice à previsão do comportamento e dos resultados do modelo (SEAVER, 2019).

O problema da inescrutabilidade e da opacidade não apenas reforça o argumento da existência de uma falsa neutralidade algorítmica, como também traz à tona o imbróglio da responsabilização jurídica:

> Quem deve ser responsabilizado quando um robô, software ou dispositivo executado por algoritmos de redes neurais artificiais discrimina pessoas ou segmentos sociais, cria acidentes ou gera resultados injustos, ofensivos e até letais? Os seus desenvolvedores, analistas e pessoas envolvidas na sua modelagem? A corporação que o comercializa? A empresa ou o governo que o adotou? E a pergunta derradeira é se podemos responsabilizar alguém por algo que tem suas decisões geradas de um modo impossível de rastrear e, portanto, de compreender e explicar (SILVEIRA, 2020, p. 91).

A responsabilização algorítmica é frequentemente encontrada na literatura específica sob o termo *accountability*, que não possui tradução exata em português pela ausência da existência do próprio conceito na sociedade, motivo pelo qual dispõe-se, no uso recorrente da linguagem, da palavra em outro idioma. Segundo Campos (1990), o conceito de *accountability* envolve, de um lado, a delegação de uma responsabilidade e, do outro lado, a gestão dos recursos delegados, gerando, com isso, uma obrigação de prestação de contas daquele que realiza a referida gestão.

O tema vem sendo mundialmente debatido, apesar de ainda não se haver chegado a uma resposta consensual para situações práticas recorrentes.

Uma das propostas mais palpáveis na atualidade é a Resolução para o Parlamento Europeu da Comissão sobre Disposições de Direito Civil sobre Robótica, aprovada em 2017. No documento, a comissão, analisando soluções jurídicas possíveis, considera a criação de um estatuto jurídico específico para robôs detentores de personalidade eletrônica (PARLAMENTO EUROPEU, 2017). Essa resposta, no entanto, foi altamente criticada por cientistas e especialistas em IA na *Open Letter To the European Commission Artificial*

Intelligence and Robotics, em tradução livre, Carta Aberta à Comissão Europeia de Inteligência Artificial e Robótica.

A carta aberta afirma que a criação de uma personalidade eletrônica para robôs autônomos, imprevisíveis e autoaprendizáveis se baseia na ideia equivocada da impossibilidade de provar responsabilização, enfatizando ainda que um estatuto jurídico dessa natureza não poderia derivar do modelo de pessoa física (já que isso possibilitaria às IAs a detenção de direitos humanos) ou do modelo de pessoa jurídica (que implicaria a existência da dirigência e controle por pessoas humanas) (ROBOTICS, s/d[12]).

A seguir, são apresentados dados acerca de utilizações de agentes inteligentes no extrativismo de dados no Brasil e no mundo em diversas áreas, havendo um enfoque para a seara jurídica.

1.2 EXTRATIVISMO ALGORÍTMICO DE DADOS

Na seara global, levantamentos sobre o uso da inteligência artificial apresentam números preocupantes.

O AI Index Report, uma iniciativa da Universidade de Stanford para rastrear, comparar e visualizar dados relativos à IA, a nível global, em seu relatório de 2023[13], aponta para um aumento exponencial de pesquisas sobre IA: os números mais que dobraram desde 2010, sendo os tópicos mais abordados o reconhecimento de padrões, o aprendizado de máquina e a visão computacional. Apesar disso, o relatório apresenta a informação de que, até o ano de 2014, os modelos de aprendizado de máquina mais significativos haviam sido lançados pela academia, mas, desde então, foi a indústria quem assumiu as rédeas: ainda em 2022, havia 32 modelos significativos de aprendizado de máquina produzidos pela indústria e apenas 3 produzidos pela academia. Isso porque construir sistemas de IA tem dependido cada vez mais de um enorme volume de dados, poder de computação e dinheiro - o GPT-2, por exemplo, lançado em 2019, tinha 1,5 bilhão de parâmetros e custou cerca de US$ 50.000 (AI INDEX, 2023).

Mas não é só isso: à medida que aumentam seus custos, aumenta a evolução de suas performances técnicas. Esses sistemas passam a gerar respostas incoerentes ou falsas altamente verossímeis, além de tornar possível a programação da IA pelo próprio agente inteligente, criando um modelo de IA de autoaperfeiçoamento. Neste ponto, discute-se a

[12] Disponível em: http://www.robotics-openletter.eu/. Acesso em 17 de maio de 2023.
[13] Disponível em: https://aiindex.stanford.edu/report/. Acesso em 18 de maio de 2023.

ética, a justiça e o viés no aprendizado de máquina: temáticas atinentes a este objeto são os modelos de *deep fake*, engendrados pelos modelos generativos, a construção de diversos sites de checagem automatizada a partir de conjuntos de bases de dados vazados e o aumento galopante dos incidentes relacionados ao uso indevido da IA, que foi 26 vezes maior do que os números de 2012. Aumentam também as aprovações de leis e casos legais relacionados à IA (AIAAIC, 2022).

O cenário não parece tão animador, mas a opinião pública é divergente: os chineses são mais positivos quanto aos usos da IA (também os homens, mais do que as mulheres, tendem a acreditar nesse potencial positivo), entre os americanos pesquisados, aqueles que relatam sentir-se entusiasmados com a IA estão mais entusiasmados com o potencial de melhorar a vida e a sociedade (31%) e economizar tempo e tornar as coisas mais eficientes (13%) (AIAAIC, 2022).

Já aqueles que relatam sentir-se mais preocupados, demonstram inquietação com a perda de empregos humanos (19%); a vigilância, o *hacking* e a privacidade digital (16%); e a falta de conexão humana (12%). Já entre os pesquisadores do Processamento de Linguagem Natural, 77% concordaram ou concordaram fracamente com a afirmação de que as empresas privadas de IA têm muita influência, 41% disseram que o PNL deveria ser regulamentada e 73% achavam que a IA poderia em breve levar a mudanças revolucionárias na sociedade; além disso, 63% acreditavam que a anonimidade do titular dos dados deveria ser preservada (AIAAIC, 2022).

A IPSOS[14], empresa de análise de dados e insights, realizou estudos recentes que demonstraram que 57% dos brasileiros acreditam que produtos e serviços que utilizam a IA apresentam mais benefícios do que prejuízos. 78% dos chineses - o país que ocupa o topo da média global - concordam com tal afirmação e apenas 35% dos estadunidenses concordam com o excerto. A média global das opiniões foi de 52%. Os países europeus do estudo (Turquia, Espanha, Rússia, Itália, Hungria, Polônia, Suíça, Bélgica, Grã-Bretanha, Alemanha, Países Baixos e França) apresentaram uma média próxima aos 44%, sendo a França o país com menor média mundial de concordância (JUNQUEIRA, 2022).

Além disso, a pesquisa apontou para o fato de que a confiabilidade na IA está correlacionada à compreensão sobre esse tipo de tecnologia - e ambos são mais elevados nos países emergentes do que nos países desenvolvidos: a China, por exemplo, está próxima aos

[14] Disponível em https://www.ipsos.com/pt-br/57-dos-brasileiros-acham-que-uso-de-inteligencia-artificial-traz-mais-vantagens-que-desvantagens. Acesso em 24 de maio de 2023.

80% de confiabilidade e aos 65% de boa compreensão do que a IA é; no Brasil, 70% da população diz ter uma boa compreensão do que a IA é e 50% de confiabilidade na tecnologia. Já os Estados Unidos e a França estão abaixo da média no que se refere tanto à confiabilidade (em torno dos 35%) quanto à compreensão de seu conceito (em torno dos 50% para a França e 65% para os Estados Unidos) (JUNQUEIRA, 2022).

Quanto às áreas mais afetadas pelo uso da IA, a amostra respondeu que, pelos próximos 3-5 anos, espera que as mais afetadas sejam educação, segurança e trabalho; e que as menos afetadas sejam nutrição, relações pessoais e direito (JUNQUEIRA, 2022).

No Brasil, mais especificamente, o estudo "Avanços na cultura organizacional baseada em dados, analytics e IA", do SAS, feito pelo IDC, revelou que o Brasil é o país que mais usa IA na América Latina, sendo os setores mais avançados o financeiro, o varejo e o de manufaturas. A pesquisa relatou que 63% das empresas aplicam IA, 90% investem em dados e 84% utilizam dados, todas as estatísticas acima das médias latinas de 47%, 60% e 73% (NOVO, 2022[15]).

Além disso, a análise identificou que as aplicações mais comuns na América Latina são vídeo em computador (48%) e automação de processos de tomada de decisão (47%), seguidas por gráficos de conhecimento (44%), reconhecimento de texto (44%), detecção de anomalia (43%), Internet das Coisas – IoT, na sigla em inglês – (42%) e reconhecimento de áudio e voz (41%) (NOVO, 2022).

A Internet das Coisas (ou, em inglês, Internet of Things - IoT) torna-se uma das preocupações da modernidade, especialmente num mundo em que o fenômeno da Internet, por si só, foi capaz de fazer com que o capitalismo nas economias de alta e média renda fosse gradativamente dominado por plataformas (SRNICEK, 2016).

A IoT é um paradigma tecnológico que diz respeito à interconexão de objetos físicos com a Internet, permitindo a troca de dados entre as tecnologias envolvidas: sensores, dispositivos de comunicação ou quaisquer outros aparelhos conectados à rede se tornam "inteligentes", capazes de captar informações e interagir com o ambiente no qual estão inseridos (LIU et al., 2013). Tal fenômeno torna-se uma preocupação também diante dos dados apresentados pelo referido estudo, que demonstram que 70% das empresas falha no que se refere à conformidade do tratamento de dados às normas de privacidade vigentes (NOVO, 2022).

---

[15] Disponível em .https://blogs.sas.com/content/sasla/2022/10/28/avancando-rumo-a-cultura-organizacional-baseada-em-dados/. Acesso em 24 de maio de 2023.

Não por acaso, a seara jurídica, como todas as outras, foi sensivelmente afetada pelo fenômeno moderno da supergeração de dados, que engendrou, para o direito, o chamado *big data* judicial: milhões de dados e metadados de sujeitos de direito acumulados em bases de dados jurídicas - o que traz um nova possibilidade para uma ciência que, por séculos, foi deôntica: analisar e predizer comportamentos (ZENCOVICHI, 2019).

No cenário jurídico global, a utilização da IA vem se valendo, principalmente, da mineração de textos, capaz de recuperar ou extrair informações de excertos, e do o Processamento de Linguagem Natural (PLN), que permite a manipulação da linguagem humana por agentes inteligentes. Tais técnicas são utilizadas para reduzir técnicas manuais de reprodução de textos jurídicos, seja a partir do gerenciamento de corpus e indexação para a leitura de conjuntos de documentos jurídicos (normas, precedentes, informações pessoais das partes) ou da sintetização, processamento ou conversão de grandes excertos (SALOMÃO, 2021).

Exemplos célebres de agentes inteligentes utilizados no cenário jurídico mundial, para além do já anteriormente mencionado COMPASS, são o ROSS Intelligence, a Kira, o LawGeex, o eBrevia, o X-Law, o *Public Safety Assessment* e o *DoNotPay*.

O Ross Intelligence, assim como a Kira, o LawGeex e o eBrevia, são ferramentas que utilizam inteligência computacional, a partir do processamento de linguagem natural, para conduzir pesquisas por meio de questionamentos em linguagem humana (HOULIHAN, 2017).

O X-Law, implementado na Itália, é uma ferramenta de combate à criminalidade que estima, a partir do perfilamento de criminosos e de dados sobre as rotinas das cidades, locais e horários mais propensos à ocorrência de delitos. A cada meia hora, o sistema envia um alerta à polícia, informando o local e a probabilidade da ocorrência de um crime nas duas horas subsequentes (SILVA, 2022).

O *Public Safety Assessment,* assim como o COMPASS, é utilizado para a análise de risco e controvérsias nos tribunais, sendo mais aplicado no Estado de New Jersey a casos de pré-julgamento (fiança e prisão cautelar)[16].

O *DoNotPay* é o primeiro agente inteligente envolvido em uma polêmica sobre a prestação de assistência jurídica em tribunal em tempo real. A IA seria utilizada na corte americana para ouvir toda a acusação e, através de um ponto eletrônico, dizer ao réu

[16] Disponível em https://www.sajdigital.com/exemplos-inteligencia-artificial. Acesso em 24 de maio de 2023.

exatamente o que falar ao juiz; essa utilização, no entanto, foi alvo de duras críticas de juristas americanos[17].

Além destes modelos, mais célebres, há exemplos menos conhecidos: na Estônia, um juiz-robô analisa disputas legais cujo valor da causa seja menor que sete mil euros; em Portugal, o Ministério da Justiça utiliza uma ferramenta desenvolvida pela Watson (IBM) para pesquisas rápidas em big datas e uma plataforma que permite a busca rápida por jurisprudências e suas interconexões com acórdãos ou processos; os Países Baixos utilizam o projeto LEDA (Legislative Design and Advisory System) para auxiliar na resolução de problemas de organização de um ato normativo, identificando, por exemplo, se determinado projeto legislativo cumpre um conjunto de requisitos para sua legitimidade (SCAPINI, 2020).

No que se refere ao cenário nacional, as etapas de digitalização da justiça (que começam com a digitalização dos processos, evoluindo para a digitalização dos procedimentos e chegando, por fim, à automação das tarefas) são marcadas pela paulatina informatização do Judiciário (SALOMÃO, 2021).

Esse processo, que teve início em 1996, com a implantação da urna eletrônica, rapidamente evoluiu para a disponibilização de serviços pela Internet, como as consultas processuais e o envio de boletins informativos, e para o peticionamento eletrônico, que teve início no mesmo ano. Grandes marcos desse processo são o ano de 2004, caracterizado pela publicação de uma série de normas que visavam regulamentar o peticionamento eletrônico, e o ano de 2021, durante o qual ocorre o lançamento do programa Justiça 4.0, cujo objetivo é ampliar a prestação jurisdicional, facilitando o acesso à justiça; tal programa engloba o Juízo 100% Digital, o Balcão Virtual, a Plataforma Digital do Poder Judiciário (PDPJ), o desenvolvimento do DataJud, a plataforma Codex e a disseminação da plataforma Sinapses[18] (SALOMÃO, 2021).

A primeira fase da pesquisa Tecnologia Aplicada à Gestão dos Conflitos no Âmbito do Poder Judiciário Brasileiro, publicada pela Fundação Getúlio Vargas (FGV) em 2021, identificou o nível de maturidade de projetos de IA presentes no judiciário brasileiro, tendo indicado que as tarefas concentravam-se na estruturação dados (fluxos de organização, triagem, automação, recuperação e extração de informações, em sua maioria). Já sua

[17] Disponível em https://legalinteract.com/robot-lawyer/. Acesso em 24 de maio de 2023.

[18] O Juízo 100% Digital retira a necessidade de comparecimento do cidadão aos fóruns, oportunizando os atos processuais digitais; o Balcão Virtual permite às varas realizar atendimento remoto; a PDPJ consolida os sistemas eletrônicos do Judiciário brasileiro em um ambiente unificado; o DataJud armazena dados e metadados relativos a processos de tribunais; e o Codex, por fim, é uma plataforma que alimenta o DataJud de forma automatizada (SALOMÃO, 2020).

segunda fase revelou que 44 tribunais, além do Conselho Nacional de Justiça (CNJ), utilizam agentes inteligentes[19] e que tal aplicação impacta mais de 75 milhões de processos em tramitação no judiciário (SALOMÃO, 2021).

Os agentes inteligentes mais conhecidos do cenário brasileiro são: Victor, utilizado pelo Supremo Tribunal Federal (STF) para identificar recursos extraordinários que possuam temas de repercussão geral; Sócrates/Athos, utilizado pelo Supremo Tribunal de Justiça (STJ) para analisar processos semelhantes que possam ter decisões idênticas; Legal Labs/Dra. Luzia, utilizada pela Procuradoria-Geral do Distrito Federal para analisar o andamento processual de execuções fiscais e sugerir possíveis soluções, além de indicar eventuais endereços e bens dos executados; Legal Labs/Victoria, utilizada para agilizar as atividades cartorárias, interpretando peticionamentos, realizando fluxos de bloqueios e gerando decisões com *deep learning* (SILVA, 2022).

O levantamento demonstrou, ainda, que a maior aplicação de agentes inteligentes no Brasil está concentrada no Centro-Oeste (mais especificamente no Distrito Federal), com 20 iniciativas, seguida pelo Sul, com 13 iniciativas, Sudeste, com 12, Nordeste, com 10, e Norte, com 9 iniciativas. Destaca-se que, do número total, apenas uma iniciativa é do Estado de Minas Gerais (SALOMÃO, 2021).

Além disso, dos projetos citados, 3% são de modelos de aprendizagem de reforço, 49% de aprendizagem supervisionada, 26% de aprendizagem não supervisionada e 23% de modelos que não utilizam a aprendizagem de máquina (SALOMÃO, 2021).

E a performance desses agentes inteligentes vem sendo tão estudada quanto suas hipóteses de aplicação.

A AIAAIC, uma iniciativa independente, não partidária e de interesse público que examina e defende a transparência e a abertura reais de IA, algoritmos e automação, possui

---

[19] Citam-se, para fins de conhecimento: Sinapses, Victor, Athos, DataJud, Indexação de Peças Processuais em Processos Originários, Identificação de Fundamentos de Inadmissão do REsp, Bem-te-vi, ALEI, Sigma, Agrupamento de apelações por similaridade de sentença, JULIA (Jurisprudência Laborada com Inteligência Artificial), I.A. de classificação de petições intermediárias, Robô Hércules, TIA, Projeto Temas Repetitivos, LEIA Precedentes, Programa Cientista Chefe, HORUS, AMON, TOTH, SAREF, ARGUS, ARGOS, Busca Eletrônica em Registros usando linguagem Natural (BERNA), OMNI, Painéis de BI para Vara da Saúde, JURIMETRIA COM IA, MIDAS, LARRY, ELIS, Módulo de classificação automática de documentos, GPSMed, Peticionamento Inteligente, MANDAMUS, IA Execução Fiscal, Chatbot DIGEP, GRAFO, Incremento dos mecanismos de pesquisa de Jurisprudência com Inteligência Artificial, Classificador de petições em Execuções Fiscais, Análise de guias, JUDI, Inteligência Artificial como ferramenta de auxílio jurisdicional, Sistema de Classificação de Petições Judiciais, Modelo de Inteligência Artificial para identificação automática de processos em trâmite na Justiça do Trabalho no sistema PJe cujo tema esteja sobrestado por determinação de órgão superior, Implantação de Data Lake, Clusterização de Processos (Recurso de Revista), Índice de Conciliabilidade por Inteligência Artificial (ICIA), GEMINI, Análise de Pressupostos de Admissibilidade, MAGUS, Seguro Garantia, Concilia JT, Sistema Automatizado de Busca Patrimonial, Sistema de BI Hórus, Sistema de Jurisprudência (SALOMÃO, 2020).

um repositório que pretende tornar transparente os processos de design, automação, desenvolvimento e implantação da IA[20]. O relatório aponta para 1.012 incidentes envolvendo o mau uso de tais modelos desde o ano de 2009, 995 são casos globais.

Analisando o documento público, é possível perceber que apenas um caso foi registrado em 2009; nenhum em 2010; um em 2011; 10 em 2012; 4 em 2013; 10 em 2014; 15 em 2015; 31 em 2016; 55 em 2017; 73 em 2018; 101 em 2019; 211 em 2020; 213 em 2021; 79 em 2022 e 55 em 2023, até o momento (AIAAIC, 2023).

Um dos incidentes notáveis em 2022 incluiu um vídeo *deepfake* do presidente ucraniano Volodymyr Zelensky se rendendo e outro de prisões dos EUA usando tecnologia de monitoramento de chamadas em seus internos (AIAAIC, 2023).

No Brasil, ainda de acordo com o documento, são citados 11 casos de uso da IA - dentre os quais 5 restritos somente ao país, denominados *"ChatGPT powers automated content, spam farms", "Microsoft teen pregnancy predictions", "Worldcoin 'field testing'", "Sao Paulo METRO SecureOS facial recognition", "Tinder Plus personalised pricing algorithm discrimination", "Facebook political ads misidentification", "São Geraldo Magela drone delivery", "Facebook Marketplace Amazon rainforest sales", "Rio de Janeiro facial recognition wrongful arrests", "Sao Paulo METRO advertising facial biometrics"* e *"President Balsonaro/Chapulin Colorado deepfake"* (AIAAIC, 2023).

O caso *ChatGPT powers automated content, spam farms* ocorreu em maio de 2023, quando o ChatGPT ou chatbots semelhantes passaram a publicar postagens de baixa qualidade e spams em vários idiomas a fim de atrair cliques, a técnica denominada fazenda de conteúdo[21] (AIAAIC, 2023).

O caso *Microsoft teen pregnancy predictions* ocorreu com a criação do controverso sistema algorítmico usado para prever a gravidez na adolescência em países latino americanos (a Plataforma Tecnológica de Intervenção Social, desenvolvida em 2016 pela Microsoft). O modelo se baseia na idade, etnia, país de origem, eventuais deficiências e dados socioeconômicos de aproximadamente 200.000 mulheres e meninas (12.000 com idades entre 10 e 19 anos) para prever com cinco ou seis anos de antecedência - com nome, sobrenome e endereço - quais das meninas tem mais de 86% de chance de ficar grávida na adolescência. A IA foi criticada por ser tendenciosa contra grupos indígenas locais e uma

[20] Disponível em: https://www.aiaaic.org/about-aiaaic/research-citations-and-mentions. Acesso em 22 de maio de 2023.

[21] As fazendas de conteúdo consistem na utilização de dados de algoritmos para a criação de mídias projetadas visando a obtenção de uma classificação elevada em mecanismos de pesquisa.

auditoria concluiu que o sistema não levava em conta o acesso à educação sexual e à informação contra concepção (AIAAIC, 2023).

O caso *Worldcoin 'field testing* remete à criptomoeda (moeda digital) homônima, acusada de enganar e explorar pessoas enquanto constrói um negócio de autenticação digital anônima: descrevendo-se como uma 'nova moeda global de propriedade coletiva baseada no Ethereum que será distribuída de forma justa para o maior número possível de pessoas', a Worldcoin usa uma esfera cromada para escanear as íris e os rostos das pessoas que concordam em se inscrever no modelo (AIAAIC, 2023).

O caso T*inder Plus personalised pricing algorithm discrimination* ocorreu com a descoberta de que os valores do plano *Tinder Plus* estavam sendo cobrados conforme características pessoais dos usuários. Dados como a localização, o gênero, a orientação sexual e, principalmente, a idade influenciavam a precificação do algoritmo (AIAAIC, 2023)..

O caso *Facebook political ads misidentification* ocorreu quando um conjunto de pesquisadores descobriu que o Facebook classificava equivocadamente 83% dos anúncios políticos veiculados na plataforma. Os pesquisadores exibiram 189.000 anúncios políticos entre julho de 2020 e fevereiro de 2021 por meio do então novo processo de autorização do Facebook para publicidade política e descobriram que a empresa identificou incorretamente 21% dos anúncios de produtos como políticos e 62% dos anúncios abertamente políticos. Também identificaram mais de 70.000 anúncios políticos veiculados no Facebook durante uma proibição temporária de publicidade política que a rede impôs durante a eleição presidencial dos EUA, muitos deles por organizações que só exibiam anúncios políticos na plataforma (AIAAIC, 2023).

O caso *São Geraldo Magela drone delivery* ocorreu quando um padre da igreja São Geraldo Magela, em Soracaba, usou um drone equipado com uma custódia para entregar a Eucaristia no altar (AIAAIC, 2023.

O caso *Facebook Marketplace Amazon rainforest sales* se pautou na descoberta, pela BBC, que as áreas da floresta amazônica brasileira, incluindo florestas nacionais e terras reservadas a povos indígenas, estão sendo compradas e vendidas ilegalmente no Facebook Marketplace (AIAAIC, 2023).

O caso *Rio de Janeiro facial recognition wrongful arrests* ocorreu quando o governo do estado do Rio de Janeiro realizou um projeto de teste de reconhecimento facial com o objetivo de identificar criminosos e “preservar a ordem pública”. Operado pela Oi, o projeto se dividia em duas fases: a primeira, limitada a Copacabana durante o carnaval de 2019, e a

segunda, no bairro do Maracanã e próximo ao Aeroporto Santos Dumont, de junho a outubro de 2019. Houve uma taxa de erro de 90% e foram realizadas várias prisões equivocadas. As autoridades declararam a primeira fase um sucesso, com cinco mandados de busca e apreensão cumpridos, três mandados de prisão emitidos e três veículos recuperados (AIAAIC, 2023).

Durante a segunda fase do projeto, a polícia do Rio prendeu onze pessoas em partidas no estádio do Maracanã. Sob pressão de O Panóptico, projeto que monitora o uso de reconhecimento facial no Brasil, as autoridades foram obrigadas a admitir que sete deles eram falsos positivos (AIAAIC, 2023).

O caso *Sao Paulo METRO advertising facial biometrics* ocorreu quando a Via Quatro, operadora da Linha Amarela do Metrô de São Paulo, instalou portas de plataforma que exibem anúncios e informações e usam sensores com telas e reconhecimento facial e de emoção para monitorar a reação dos telespectadores. O movimento resultou em defensores dos direitos humanos e da privacidade para expressar preocupações sobre a imprecisão dos sistemas biométricos faciais, o potencial de preconceito racial e étnico e de pseudociência.

O desdobramento desse caso foi o *Sao Paulo METRO SecureOS facial recognition,* quando a juíza Cynthia Thome, do tribunal de São Paulo, determinou que a empresa responsável pela administração do metrô de São Paulo suspendesse o uso de seu sistema de reconhecimento facial como parte de sua implementação mais ampla do sistema de vigilância eletrônica SecurOS. A Companhia do Metropolitano de São Paulo (METRO) foi demandada em virtude de abuso de privacidade e por não ter informado ou recebido consentimento dos usuários do serviço no que se refere ao objetivo e escopo de seu sistema de biometria facial (AIAAIC, 2023).

O caso *President Balsonaro/Chapulin Colorado deepfake*, por fim, consistia na criação de mídias virais do ex-presidente Bolsonaro em diversos contextos (AIAAIC, 2023).

1.2.1 Análise da dataficação-financeirização do extrativismo digital

Se antes já havia o entendimento de que o capital pode ser cultural, econômico e social (BORDIEU, 2007), o processamento de dados no centro de um novo sistema econômico passa, obrigatoriamente, pela compreensão de uma mudança de paradigma: o fluxo de renda intermediado por ativos agora diz respeito a novas formas de financeirização que compreendem todo e qualquer tipo de capital.

A modernização do capital parece, portanto, encontrar bases firmes no mercado de exploração de dados, o que pode configurar o motivo pelo qual a indústria assumiu as rédeas da produção e manipulação dos modelos de aprendizado de máquina mais significativos.

O Brasil encabeça estatísticas de empresas que aplicam em IA, investem em dados e utilizam dados e, apesar de o potencial de extrativismo desse recurso infinito gerar uma preocupação global, especialmente em relação aos sistemas inteligentes envolvidos na operação, como já discutido anteriormente, a maioria dos brasileiros acredita que a utilização da IA apresenta mais benefícios do que prejuízos, confiando nos modelos inteligentes e afirmando ter boa compreensão do que a tecnologia é.

Ainda assim, o que vem se comprovando, tanto no cenário global quanto no cenário nacional, é que os incidentes envolvendo mau uso de modelos inteligentes aumenta exponencialmente.

Outra previsão brasileira que parece não se cumprir é a de que o Direito esteja entre as três áreas menos afetadas pelo uso da IA nos próximos 3 a 5 anos: o que se percebe é que a digitalização do judiciário vem ocorrendo desde 1996 e, com o surgimento do *big data* judicial, a expectativa é de que os agentes inteligentes estejam cada vez mais presentes. Além disso, a segunda aplicação mais comum da IA na América Latina é a automação de processos de tomada de decisão, o que pode ser de grande valia para o Direito.

A utilização da inteligência artificial pelo próprio Direito evidencia a necessidade de se compreender de que forma os juristas vêm enfrentando o tema. Por esse motivo, a seguir, apresenta-se o estado da arte da legislação sobre o tema, a fim de se averiguar a construção teórica sobre o assunto.

2 ESTADO DA ARTE NO CENÁRIO JURÍDICO

Este capítulo trata da evolução normativa e do estado da arte no cenário jurídico nacional e internacional. Nessa seara, um dos aspectos mais latentes da inovação tecnológica é sua dificuldade regulatória, mais especialmente nos estágios iniciais do desenvolvimento da tecnologia, ante a falta de informações sobre seu impacto; por outro lado, em um estágio mais avançado, a tecnologia se torna mais difusa, tornando a força regulatória mais dispendiosa e menos efetiva.

Dessa forma, se, num primeiro momento, a conclusão é de que o Estado deveria aguardar a maturidade tecnológica para implementar a regulação, gerando uma insegurança jurídica temporária; num momento posterior, no qual a tecnologia já estaria plenamente desenvolvida e absorvida pela sociedade, a tentativa regulatória pode ter um custo impeditivo muito mais alto, podendo, inclusive, se tornar inócua, por estar mais solidamente ligada à vida social (MOSES, 2014).

Esse paradoxo, relacionado à dificuldade temporal de se regular uma nova tecnologia e denominado dilema de Collingridge (SAIKALI, 2020), é o que desafia os quadros jurídicos clássicos e nos coloca diante de uma revolução tecnológica e, por consequência, jurídica (SCALPINI, 2020 *apud* SOARES PEREIRA; LOPES ROCHA, 2020).

A despeito da incerteza regulatória e informacional que orbita o tema da inteligência artificial, certo é que os reguladores, pautando-se ou não em instrumentos normativos existentes, são forçados a tomar decisões (SAIKALI, 2020). O objetivo deste capítulo é perpassar a dicção normativa existente a respeito da utilização da IA e, posteriormente, averiguar se tal lógica regulatória se mantém no momento de sua aplicação.

2.1 CENÁRIO INTERNACIONAL

Como já apresentado anteriormente, a problemática do manejo indevido dos dados pessoais no ciberambiente moderno e a cleptocracia digital que lidera a guerra de dados da quarta revolução industrial trazem à tona a necessidade constante de reinvenção do direito, a fim de abarcar as questões tecnológicas advindas da nova era contemporânea.

Assim, algumas normativas foram criadas, no cenário nacional e internacional, na tentativa de acompanhar o desenvolvimento desenfreado das TICs e da consequente

evolução de formas de extração, mineração e tratamento de dados pessoais. Citam-se, a seguir, aquelas que foram consideradas as mais relevantes para o fim deste estudo.

Inicialmente, cabe destacar que o Brasil possui, ainda, um número limitado de normativas que regulam o ciberambiente, principalmente porque uma das primeiras leis a tratar do tema, a Lei nº 12.965, denominada Marco Civil da Internet, foi promulgada em 2014, há apenas 8 (oito) anos.

O cenário internacional, por outro lado, já caminha a passos mais largos nessa literatura.

Na América do Norte, o Canadá possui 28 regulamentações, entre leis provinciais e federais, sobre a proteção de dados. A Lei de Proteção de Informações Pessoais e Documentos Eletrônicos (Personal Information Protection and Electronic Documents Act - PIPEDA), de 2000, opera sobre princípios de boas práticas - como o princípio da necessidade de identificação dos propósitos por trás de uma coleta - e contém diretrizes que vinculam a coleta, o tratamento e a divulgação dos dados pessoais. Os Estados Unidos não possuem uma única lei para a proteção de dados, mas as normativas mais importantes, como a Lei de Privacidade de 1974, a Lei de Proteção à Privacidade de 1980, a Lei Gramm-Leach-Bliley de 1999, a Lei de Portabilidade e Responsabilidade dos Seguros de Saúde de 1999 e a Lei de Relatório de Crédito Justo de 2018, demonstram a preocupação legislativa com o tema (GONZÁLEZ, 2020).

Na América Central, o México possui normativas como a Lei Federal de Proteção de Dados Pessoais em Poder de Particulares, desde o ano de 2010, e a lista de Regulamentações da Lei Federal de Proteção de Dados Pessoais em Poder de Particulares, de 2011, além do Instituto Federal de Acesso à Informação e Proteção de Dados (IFAI) (GONZÁLEZ, 2020).

Na América do Sul, a Argentina possui a Lei nº 25.326, Lei de Proteção de Dados (PDPA), sancionada desde o ano de 2000, estabelecendo coleta de dados mediante consentimento e direito de acesso, correção e exclusão pelo titular. Na Colômbia, vigem a Lei 1.266/08, a Lei 1.273/09, a Lei 1.581/12 e o Decreto 1.377/13, que abarcam desde o consentimento do titular, passando pelo processamento e transferência de dados, até o cometimento de cibercrimes - prevendo como delito o roubo, compra e venda de dados pessoais (GONZÁLEZ, 2020).

Na Ásia, a Índia não tem uma lei central sobre a proteção de dados, mas publicou uma normativa em dezembro de 2019 que está em análise pela comissão parlamentar; até a aprovação, as normativas mais importantes são a Lei de Tecnologia da Informação (Lei 21

de 2000) e a lista de Regras de Tecnologia da Informação de 2011. A China possui atualmente a Tecnologia da Informação: Especificação Sobre Segurança de Informações Pessoais (GB/T 35273-2017) - conhecida como “O Padrão” - e possuía, antes disso e desde 2014, um conjunto de regras miscelâneas que abordavam o tema. O Japão possuía desde 2003 a Lei 57 que tratava da privacidade de dados pessoais e promulgou, em 2017, a Emenda APPI, que traz novas regras sobre compartilhamento de dados com terceiros, anonimização e vazamentos (GONZÁLEZ, 2020).

Na Oceania, a Austrália possui a Lei de Privacidade de 1988, construída sobre os 13 Princípios Australianos de Privacidade (APPs, ou *Australian Privacy Principles*), e que aborda assuntos que orbitam o uso e divulgação de dados, direitos do titular, transparência e anonimidade, além da manutenção da qualidade das informações. A Nova Zelândia tem um modelo parecido: segue os 12 Princípios da Privacidade de Informação, estabelecidos em 1993 na Lei de Privacidade do país e possui legislações específicas voltadas para setores como o de crédito, o de saúde e o de telecomunicações (GONZÁLEZ, 2020).

Na Europa, o Regulamento Geral sobre a Proteção de Dados 2016/679 (GDPR), criado em 2018, é tido como um exemplo mundial de uma normativa de proteção de dados a ser seguido. Não por acaso, a Alemanha, um dos países líderes na regulamentação sobre privacidade e proteção de dados e que já possuía uma normativa vigente desde 2001 (a Bundesdatenschutzgesetz – BDSG[22]), agora possui a Lei Federal de Proteção de Dados de 2017 (que revogou a BDSG), mais alinhada à GDPR. Países como a França, a Grécia e a Dinamarca seguiram a mesma trajetória (GONZÁLEZ, 2020).

A GDPR não é uma norma voltada especificamente para a regulação da IA, mas estabelece regras estritas para o tratamento, coleta, processamento e uso de dados pessoais que podem também ser aplicados em agentes inteligentes. Além disso, existem também as regulações específicas para esses modelos.

Em 2016, foi publicado o Relatório do Parlamento Europeu que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica e ao qual seguiu sua Resolução, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica de 2017.

O Parlamento Europeu propôs também uma resolução que recomendava a criação da denominada Lei de Responsabilidade Civil para Robôs, que regularia a atribuição de

[22] A BDSG trata dos direitos e deveres de órgãos públicos e privados para as atividades de coleta e processamento de dados, que têm o dever de contratar um profissional responsável por privacidade de dados e de determinar regras claras para avaliações de score de crédito, por exemplo. Além disso, há diretrizes específicas para como as empresas devem e podem fazer tratamentos de dados de seus funcionários.

responsabilidade por danos causados por sistemas autônomos. A Resolução 2015/2103-INL, editada em 2017 pelo Parlamento Europeu, apresenta, inicialmente, as características necessárias para determinado modelo ser considerado inteligente: a) existência de sensores capazes de permitir a troca de dados com o ambiente; b) capacidade de aprendizado com a experiência e interação com o meio; c) existência de um suporte material; d) capacidade de adaptação; e e) ausência de vida na acepção biológica. Dessa forma, à medida que a autonomia de determinado modelo aumenta, não é possível considerá-lo apenas uma ferramenta nas mãos de outro agente, o que levaria também a uma discussão sobre a personificação de sistemas autônomos (EUROPA, 2017).

A inteligência dessa normativa foi idealizada a partir da consideração de IAs cada vez mais sofisticadas e preparadas para desencadear uma revolução industrial e da tendência do desenvolvimento de tais agentes inteligentes para a tomada de decisões; a proposta também é pautada nos seguintes princípios: transparência, beneficência, não-maleficência, autonomia e justiça, além de direitos fundamentais tais como a dignidade do ser humano, a igualdade, a justiça e a equidade, a não discriminação, o consentimento esclarecido, o respeito pela vida privada e familiar e a proteção de dados, bem como em outros princípios e valores subjacentes do direito da União, como a não estigmatização, a transparência, a autonomia, a responsabilidade individual e a responsabilidade social, e em códigos e práticas éticas existentes, tendo em vista que deve ser dada particular atenção aos robôs que constituem uma ameaça importante à privacidade devido ao seu posicionamento em espaços tradicionalmente protegidos e íntimos e à sua capacidade de extrair informações relativas a dados pessoais sensíveis e de os transmitir (EUROPA, 2017).

Para além das normativas europeias, há que se considerar as comunicações do Parlamento Europeu, que oferecem informações à sociedade acerca de questões de interesse para a UE, nomeadamente as políticas e as ações da União.

Nesse ínterim, serão analisadas mais detalhadamente as comunicações que se referem direta ou indiretamente à inteligência artificial; são elas: “Inteligência artificial para a Europa”, de 2018; “Plano Coordenado para a Inteligência Artificial”, também de 2018; e “Aumentar a confiança numa inteligência artificial centrada no ser humano”, de 2019. Outras serão citadas a seguir, conforme o critério cronológico.

Também no ano de 2018 cabe destacar as Diretrizes Universais para Inteligência Artificial[23], publicadas pela The Public Voice, uma ONG que tem por objetivo promover a participação pública da sociedade civil especialmente em assuntos que envolvam a liberdade de expressão e o direito à privacidade; e a Declaração sobre Ética e Proteção de Dados em Inteligência Artificial[24], uma contribuição da América Latina e do Caribe ao debate homônimo, publicada pelo *International Conference of Data Protection and Privacy Commissioners* (ICDPPC), um fórum anual que trata sobre a regulação de privacidade.

Porque o uso da IA pelo Judiciário é tido, no imaginário social, como um critério de qualidade e efetividade do sistema de Justiça, a Comissão Europeia para a Eficácia da Justiça (CEPEJ) criou um grupo de trabalho (GT) sobre a qualidade da justiça responsável por orientar a aplicação dessas ferramentas nos tribunais europeus. Em dezembro de 2018, o trabalho deste GT resultou na Carta Ética Europeia sobre o uso de IA nos sistemas judiciais (EUROPA, 2018).

Em outubro de 2019, foi realizada uma conferência organizada no âmbito do Comitê de Ministros do Conselho da Europa, com o tema "Justiça na Europa face aos desafios da Era digital" e, em dezembro de 2019, a CEPEJ criou um novo GT sobre ciberjustiça e IA para aprofundar-se no assunto, com o objetivo de oferecer uma normativa que oriente e preveja garantias aos Estados-membros e aos profissionais que pretendam criar ou utilizar tecnologias e/ou modelos de IA nos sistemas judiciais para melhorar a eficiência e a qualidade da Justiça (FGV, 2021). Essa ação culminou no lançamento do plano de ação "Digitalização para uma justiça melhor" (*Digitalisation for a better justice*) para o período de 2022 a 2025, cujo objetivo é acompanhar os Estados e os tribunais numa transição bem-sucedida para a digitalização, em conformidade com as normas europeias, em particular o artigo 6.º da Convenção Europeia dos Direitos do Homem.

Também no ano de 2019, cabe ressaltar a publicação, pela Comissão Europeia, do Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica[25].

Em 19 de fevereiro de 2020, a Comissão apresentou o Livro Branco sobre a inteligência artificial. O documento antecipa quadros regulamentares para uma inteligência

[23] Disponível em: https://thepublicvoice.org/ai-universal-guidelines/. Acesso em 8 de jul. de 2023.

[24] Disponível em: https://www.alsur.lat/sites/default/files/2020-05/Intelig%C3%AAncia%20Artificial_Consulta%20P%C3%BAblica_%20ICDPPC_AI_Sur.pdf. Acesso em 8 de jul. de 2023.

[25] Disponível em: https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_ES.html. Acesso em 8 de jul. de 2023.

artificial fiável, a fim de analisar uma série de riscos que devem ser discutidos e, principalmente, regulamentados, garantindo os melhores resultados a nível social, ambiental e econômico, além da conformidade com a legislação e princípios da União Europeia. A inteligência dessa norma demonstra sua preocupação com a utilização negativa da IA, seja ela material, referindo-se à segurança e saúde de pessoas, ou imaterial, referindo-se à perda de privacidade, a limitações ao direito à liberdade de expressão ou da dignidade humana e a discriminações de quaisquer tipos. O objetivo da proposta é minimizar potenciais riscos advindos dessa má utilização, em especial os mais significativos, no geral relacionados à proteção de garantias fundamentais (EUROPA, 2020).

O Relatório sobre as implicações em matéria de segurança e de responsabilidade decorrentes da inteligência artificial, da Internet das coisas e da robótica[26], que acompanha o Livro Branco, vem reforçar a existência de uma lacuna no atual quadro no que diz respeito aos riscos específicos engendrados pelos sistemas de IA e outras tecnologias digitais (SCAPINI, 2020).

Já o Relatório que contém recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial, que também acompanha o livro branco, considera que os sistemas de IA representam importantes desafios jurídicos para o quadro de responsabilidade civil em vigor e defende o mesmo nível de proteção que existe nos casos de danos causados por humanos para os casos em que existem danos causados por sistemas de IA, sustentando a ideia de uma legislação uniforme para União Europeia (SCAPINI, 2020).

Em anexo ao relatório, e tendo em conta o artigo 114º TFUE, surge a proposta de regulamento do Parlamento Europeu e do Conselho relativa à responsabilidade pelo funcionamento de sistemas de IA, que considera que não é necessário rever completamente os regimes de responsabilidade civil ou atribuir aos sistemas de IA personalidade jurídica; que a Diretiva da Responsabilidade dos Produtos (DPR) garante indenização pelos danos causados por produto defeituoso, podendo tal legislação ser aplicada aos modelos inteligentes; e que os atuais regimes de responsabilidade culposa proporcionam a proteção suficiente para pessoas que sofrerem danos causados por terceiro interveniente. Por fim, a proposta prevê a criação de um regulamento relativo à responsabilidade pelo funcionamento

[26] Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52020DC0064. Acesso em 8 de jul. de 2023.

de sistemas de IA, no qual os operadores respondem solidária e proporcionalmente pelos danos causados pelo modelo (SCAPINI, 2020).

Junto a este documento, é possível extrair distintas Comunicações do Parlamento Europeu que abordam diretamente questões relacionadas à IA, como “Uma estratégia europeia para os dados”; “Construir o futuro digital da Europa”, ou “Uma nova estratégia industrial para a Europa”. Tais documentos serão melhor analisados nas próximas seções.

Já em 2020, foram publicadas as resoluções sobre processos automatizados de tomada de decisão: assegurar a proteção do consumidor e a livre circulação de bens e serviços e sobre uma nova estratégia industrial para a Europa, as Conclusões do Conselho sobre a construção do futuro digital da Europa em 2020[27] e a Resolução do Parlamento Europeu que contém recomendações à Comissão sobre o regime relativo aos aspectos éticos da inteligência artificial, da robótica e das tecnologias conexas[28] (MORENO, 2021).

A Comissão Europeia apresentou também, em abril de 2021, a Proposta de Regulamento sobre Inteligência Artificial, com os objetivos específicos de:

> [...] a) garantir que os sistemas de IA colocados no mercado da União e utilizados sejam seguros e respeitem a legislação em vigor em matéria de direitos fundamentais e valores da União, b) garantir a segurança jurídica para facilitar os investimentos e a inovação no domínio da IA, c) melhorar a governação e a aplicação efetiva da legislação em vigor em matéria de direitos fundamentais e dos requisitos de segurança aplicáveis aos sistemas de IA, d) facilitar o desenvolvimento de um mercado único para as aplicações de IA legítimas, seguras e de confiança e evitar a fragmentação do mercado (EUROPA, 2021).

Visando, assim, estabelecer um quadro regulatório harmonizado para a IA em toda a União Europeia (UE), garantindo sua segurança, ética e conformidade com os direitos fundamentais.

Essa proposta de regulamento atribui três categorias de risco aos modelos inteligentes: inaceitável, alto risco e risco baixo ou inexistente, modelando o conteúdo das normas a partir do risco oferecido pelo agente inteligente (EUROPA, 2021).

Dessa forma, a responsabilidade deveria ser proporcional ao nível efetivo de instruções dadas ao robô e ao nível de sua autonomia: quanto maior a capacidade de

---

[27] Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52020XG0616%2801%29. Acesso em 8 de jul. de 2023.

[28] Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_PT.html. Acesso em 8 de jul. de 2023.

aprendizagem e/ou autonomia de um robô e quanto mais longo o seu treinamento, maior deve ser a responsabilidade do programador (SCAPINI, 2020).

Para tanto, a resolução apresenta as seguintes possíveis soluções jurídicas:

> a) Criar um regime de seguros obrigatórios, se tal for pertinente e necessário para categorias específicas de robôs, em que, tal como acontece já com os carros, os produtores ou os proprietários de robôs sejam obrigados a subscrever um seguro para cobrir os danos potencialmente causados pelos seus robôs; b) Garantir que os fundos de compensação não sirvam apenas para garantir uma compensação no caso de os danos causados por um robô não serem abrangidos por um seguro; c) Permitir que o fabricante, o programador, o proprietário ou o utilizador beneficiem de responsabilidade limitada se contribuírem para um fundo de compensação ou se subscreverem conjuntamente um seguro para garantir a indemnização quando o dano for causado por um robô; d) Decidir quanto à criação de um fundo geral para todos os robôs autónomos inteligentes ou quanto à criação de um fundo individual para toda e qualquer categoria de robôs e quanto à contribuição que deve ser paga a título de taxa pontual no momento em que se coloca o robô no mercado ou quanto ao pagamento de contribuições periódicas durante o tempo de vida do robô; e) Garantir que a ligação entre um robô e o seu fundo seja patente pelo número de registo individual constante de um registo específico da União que permita que qualquer pessoa que interaja com o robô seja informada da natureza do fundo, dos limites da respetiva responsabilidade em caso de danos patrimoniais, dos nomes e dos cargos dos contribuidores e de todas as outras informações relevantes; f) Criar um estatuto jurídico específico para os robôs a longo prazo, de modo a que, pelo menos, os robôs autónomos mais sofisticados possam ser determinados como detentores do estatuto de pessoas eletrónicas responsáveis por sanar quaisquer danos que possam causar e, eventualmente, aplicar a personalidade eletrónica a casos em que os robôs tomam decisões autónomas ou em que interagem por qualquer outro modo com terceiros de forma independente (SCAPINI, 2020).

A União Europeia promulgou, ainda, no dia 5 de julho de 2022, a Lei de Serviços Digitais - através do Ato de Mercados Digitais (DMA) e do Ato de Serviços Digitais (DSA) - para regular as denominadas “big techs”. A expectativa é de que, como a GDPR, tais atos normativos se tornem novos parâmetros regulatórios mundiais (BUTCHER, 2022). No Brasil, é possível que a tendência se confirme, já que, de maneira geral, a legislação nacional vem se pautando nas diretrizes europeias, é o caso da Resolução 332/2020 do Conselho Nacional de Justiça (CNJ), que incorporou os princípios estipulados na Carta Europeia do Comissão Europeia Pela Eficiência da Justiça (CEPEJ) e da LGPD, inspirada na GDPR (EUROPA, 2022).

É possível perceber que as instituições europeias, no geral, destacam a grande importância do componente ético e dos direitos humanos; dessa forma, conforme destacado

pelo Grupo de Especialistas de Alto Nível em Inteligência Artificial da UE[29], os requisitos essenciais para a regulamentação e gestão legal da IA, poderiam ser sintetizados pelas seguintes diretrizes: a) Intervenção e supervisão humana; b) Robustez técnica e segurança; c) Privacidade e gerenciamento de dados; d) Transparência; e) Diversidade, não discriminação e equidade; f) Bem-estar social e ambiental; g) Prestação de contas (MORENO, 2021). Dos relatórios produzidos por este grupo serão analisados em especial "Uma definição de IA: Principais capacidades e disciplinas científicas" e "Recomendações de políticas e investimentos para uma IA confiável", ambos do ano de 2019.

O protagonismo regulatório europeu no que se refere às normas de utilização da IA tem exercido influência não só sobre diversos países, mas também sobre organismos internacionais, como é o caso da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que publicou, em 2019, as Recomendações do Conselho da OCDE sobre Inteligência Artificial. As diretrizes oferecem um conjunto de princípios para o desenvolvimento e governança responsável da IA, dentre os quais é possível mencionar: transparência, responsabilidade, segurança e privacidade e seu objetivo é de que, no âmbito nacional, os Estados produzam normativas que mantenham tais preceitos. O Brasil, assim como a Argentina, o Peru, a Colômbia e a Costa Rica, além de outros 37 países membros da OCDE, concordaram em seguir tais princípios internamente (OCDE, 2019).

A OCDE recomenda, ainda, que os países signatários promovam e implementem cinco princípios para a administração responsável de Inteligência Artificial confiável. São eles: (i) crescimento inclusivo, desenvolvimento sustentável e bem-estar: a IA deve ser concebida para colocar os interesses das pessoas e do planeta em primeiro lugar, por meio do aumento das capacidades humanas, da promoção do crescimento inclusivo e do desenvolvimento sustentável, bem como da proteção dos ambientes naturais; (ii) valores e justiça centrados no ser humano: os atores de IA devem respeitar o Estado de Direito, os direitos humanos e os valores democráticos, durante todo o ciclo de vida do sistema, incluindo liberdade, dignidade e autonomia, privacidade e proteção de dados, não discriminação e igualdade, diversidade, justiça, justiça social e direitos trabalhistas internacionalmente reconhecidos. (iii) transparência e explicabilidade: sistemas de IA devem ser regidos pela transparência e divulgação responsável. Para tanto, os atores devem fornecer informações com significado, apropriadas ao contexto e consistentes com o

[29] Disponível em: Os documentos e relatórios produzidos pelo Grupo estão disponíveis em: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence. Acesso em 8 de jul. de 2023.

avanço tecnológico. Tal princípio é essencial para promover informação ampla e de fácil compreensão pela sociedade sobre os sistemas de IA, para possibilitar que os adversamente afetados desafiem seus resultados, bem como para conscientizar as partes interessadas sobre suas interações com os sistemas de IA, inclusive no local de trabalho. (iv) robustez, segurança e proteção: um sistema de IA robusto é aquele protegido durante todo o seu ciclo de vida, para que possa funcionar de forma apropriada e não coloque riscos de segurança inaceitáveis. Por isso, os atores de IA devem assegurar a rastreabilidade, incluindo a relativa aos conjuntos de dados, processos e decisões tomadas ao longo do ciclo de vida do sistema de IA. Ainda, os atores da IA devem aplicar uma abordagem contínua de avaliação e de gerenciamento de riscos. (v) responsabilidade ou prestação de contas: os atores da IA devem ser responsabilizados pelo seu funcionamento adequado e em harmonia com os princípios acima, conforme seus papéis, o contexto e consistência com o avanço tecnológico (OCDE, 2019).

Além da recomendação da OCDE e dos Princípios de Asilomar (já anteriormente mencionados), existem outras diretrizes internacionais sobre a temática. É o caso, por exemplo, da Declaração de Montreal para a Responsabilidade Artificial, publicada em 2018, destaca os princípios do bem-estar, do respeito à autonomia, de proteção da intimidade e da vida privada, da solidariedade, da participação democrática, da equidade, da inclusão da diversidade, da prudência, da responsabilidade e do desenvolvimento sustentável (MONTREAL, 2018).

Para além desse documento, existem muitos outros focados na transformação digital e na criação de uma IA ética e segura, com uma abordagem centrada no ser humano. É possível citar a Declaração Ministerial sobre Comércio e Economia Digital do G20 e a Declaração de Toronto: Protegendo os Direitos à Igualdade e à Não-Discriminação em Sistemas de Aprendizado por Máquinas (EBIA, 2021).

Este estudo não tem o intuito de exaurir as regulamentações internacionais existentes sobre o assunto, senão somente as que mais impactam o cenário nacional, que será analisado a seguir.

2.2 CENÁRIO NACIONAL

Diante do pano de fundo do cenário internacional, fica mais evidente que a legislação nacional, tanto no que se refere à proteção de dados quanto ao que se refere à regulação da inteligência artificial, ainda é comparativamente incipiente.

Apesar de o art. 5º, inciso X, da Carta Magna estabelecer o direito à privacidade desde 1988, foi somente em 2014 que foi sancionada a primeira lei que regulava o ciberambiente no Brasil.

A Lei Carolina Dieckmann, Lei nº 12.737 de 2012, tipifica crimes digitais, mencionando também delitos de invasão de dispositivos informáticos a fim de obtenção de dados (BRASIL, 2012).

O Marco Civil da Internet traz diretrizes gerais para o uso da rede e menciona muito brevemente a retenção de dados, em seu art. 15 (BRASIL, 2014).

O Decreto 8.771, de 11 de maio de 2016, que regulamenta o Marco Civil da Internet, indica procedimentos para guarda e proteção de dados por provedores de conexão e de aplicações, aponta medidas de transparência na requisição de dados cadastrais pela administração pública e estabelece parâmetros para fiscalização e apuração de infrações; já o Decreto 8.777, de 11 de maio de 2016, institui a Política de Dados Abertos do Governo Federal (BRASIL, 2016). Bases de dados abertos podem servir para a alimentação de sistemas de Inteligência Artificial, o que destaca a importância de tal normativa para o presente estudo; além disso, foi publicada a Portaria nº 46/2016, que dispõe sobre o Software Público Brasileiro.

A Lei Geral de Proteção de Dados Pessoais, Lei nº 13.709, de 14 de agosto de 2018, altera o Marco Civil da Internet para dispor sobre a proteção de dados pessoais e a Medida Provisória 869 de 2018, que altera a retrocitada lei, dispondo ainda sobre a proteção de dados pessoais e criando a Autoridade Nacional de Proteção de Dados (ANPD), que é aceita como membro da Global Privacy Enforcement Network (GPEN) (BRASIL, 2018). Foi também em 2018 que foi publicada a Instrução Normativa STJ/GP n° 6, de 12 de junho de 2018, que institui projeto-piloto de aplicação de soluções de inteligência artificial no Superior Tribunal de Justiça (BRASIL, 2018).

O processo de evolução legislativa brasileira, no que se refere ao uso da tecnologia na seara jurídica, foi preponderantemente marcado pelo Conselho Nacional de Justiça (CNJ), especialmente porque o incentivo ao acesso à justiça digital integrou um dos eixos da gestão do Ministro Fux. A primeira aproximação do CNJ com o tema foi a Resolução nº 261 de 11 de agosto de 2018, que criava o Sistema de Solução Digital da Dívida Ativa,

com objetivo de melhorar a composição entre o contribuinte e as Fazendas Públicas (BRASIL, 2018).

Posteriormente, foi publicada a Portaria nº 25 do CNJ, de 19 de fevereiro de 2019 (posteriormente revogada pela Resolução nº 395 de 07 de junho de 2021), que instituía o Laboratório de Inovação para o Processo Judicial em meio Eletrônico (Inova PJe) e o Centro de Inteligência Artificial aplicada ao PJe.

Ainda no ano de 2019 foram publicadas as resoluções nº 305, de 17/12/2019, que estabelece os parâmetros para o uso das redes sociais pelos membros do Poder Judiciário; a Portaria nº 197 de 22/11/2019, que institui um Grupo de Trabalho destinado à elaboração de estudos e propostas voltadas à ética na produção e uso da Inteligência Artificial no poder judiciário e dá outras providências.

No ano de 2019, a IA passa a integrar expressamente as atribuições da Secretaria de Tecnologias Aplicadas e de seu Departamento de Tecnologias Estratégicas e Produção, órgão incumbido de formular propostas para a implementação de políticas de IA, conforme artigos 26 e 27 do Decreto 9.677/2019 ((ENGELMANN, 2021).

Nesse mesmo ano, a Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.853, de 8 de julho de 2019, passa a dispor sobre a proteção de dados pessoais e da criação da Autoridade Nacional de Proteção de Dados e altera a Lei nº 13.709, de 14 de agosto de 2018 (BRASIL, 2019). Não por acaso, no ano de 2020 foi publicado o Decreto n° 10.222, que aprova a Estratégia Nacional de Segurança Cibernética.

No ano de 2020, foram publicadas pelo CNJ as Portaria nº 7 de 16/01/2020, que institui o Repositório Nacional de Projetos e Versionamento de Arquivos do Conselho Nacional de Justiça– Git.jus, como sistema de acompanhamento de projetos, controle de versão de arquivos e ambiente digital central para colaboração e inovação do Poder Judiciário; a Portaria nº 271 de 04/12/2020, que Regulamenta o uso de Inteligência Artificial no âmbito do Poder Judiciário; a Recomendação nº 70, de 04/08/2020, que recomenda aos tribunais brasileiros a regulamentação da forma de atendimento virtual aos advogados, procuradores, defensores públicos, membros do Ministério Público e da Polícia Judiciária e das parte no exercício do seu Jus Postulandi (art. 103 do NCPC), no período da pandemia da Covid-19; a Recomendação nº 83, de 16/12/2020, que recomenda aos tribunais brasileiros o estabelecimento de critérios para a realização de audiências, avaliação da equipe interprofissional, participação em programa e/ou curso de preparação para adoção e outros atos processuais por meio de videoconferência, enquanto perdurar o estado de calamidade pública, reconhecido pelo Decreto Federal nº 06/2020, em razão da

pandemia mundial por Covid-19; a Resolução nº 317, de 30/04/2020, que dispõe sobre a realização de perícias em meios eletrônicos ou virtuais em ações em que se discutem benefícios previdenciários por incapacidade ou assistenciais, enquanto durarem os efeitos da crise ocasionada pela pandemia do novo Coronavírus, e dá outras providências; a Resolução nº 320, de 15/05/2020, que altera a Resolução CNJ nº 185/2013, que institui o Sistema Processo Judicial Eletrônico – PJe como sistema de processamento de informações e prática de atos processuais e estabelece os parâmetros para sua implementação e funcionamento; a Resolução nº 330, de 26/08/2020, que regulamenta e estabelece critérios para a realização de audiências e outros atos processuais por videoconferência, em processos de apuração de atos infracionais e de execução de medidas socioeducativas, durante o estado de calamidade pública, reconhecido pelo Decreto Federal nº 06/2020, em razão da pandemia mundial por Covid-19; a Resolução nº 331, de 20/08/2020, que institui a Base Nacional de Dados do Poder Judiciário – DataJud como fonte primária de dados do Sistema de Estatística do Poder Judiciário – SIESPJ para os tribunais indicados nos incisos II a VII do art. 92 da Constituição Federal; a Resolução nº 332, de 21/08/2020, que dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências; a Resolução nº 334, de 21/09/2020, que institui o Comitê Consultivo de Dados Abertos e Proteção de Dados no âmbito do Poder Judiciário; a Resolução nº 335, de 29/09/2020, que institui política pública para a governança e a gestão de processo judicial eletrônico, integra os tribunais do país com a criação da Plataforma Digital do Poder Judiciário Brasileiro – PDPJ-Br, mantém o sistema PJe como sistema de Processo Eletrônico prioritário do Conselho Nacional de Justiça; a Resolução nº 337, de 29/09/2020, que dispõe sobre a utilização de sistemas de videoconferência no Poder Judiciário; a Resolução nº 341, de 07/10/2020, que determina aos tribunais brasileiros a disponibilização de salas para depoimentos em audiências por sistema de videoconferência, a fim de evitar o contágio pela Covid-19; a Resolução nº 342, de 09/09/2020, que institui e regulamenta o Banco Nacional de Medidas Protetivas de Urgência – BNMPU, nos termos do parágrafo único do artigo 38-A da Lei nº 11.340/2006, com redação dada pela Lei nº 13.827/2019; a Resolução nº 345, de 09/10/2020, que dispõe sobre o "Juízo 100% Digital" e dá outras providências; a Resolução nº 354, de 9/11/2020, que dispõe sobre o cumprimento digital de ato processual e de ordem judicial e dá outras providências; a Resolução nº 357, de 26/11/2020, que dispõe sobre a realização de audiências de custódia por videoconferência quando não for possível a realização, em 24 horas, de forma presencial; a Resolução nº 358, de 02/12/2020, que regulamenta a criação

de soluções tecnológicas para a resolução de conflitos pelo Poder Judiciário por meio da conciliação e mediação; a Resolução nº 360, de 17/12/2020, que determina a adoção do Protocolo de Gerenciamento de Crises Cibernéticas no âmbito do Poder Judiciário (PGCC/PJ); a Resolução nº 361, de 17/12/2020, que determina a adoção de Protocolo de Prevenção a Incidentes Cibernéticos no âmbito do Poder Judiciário (PPICiber/PJ); a Resolução nº 362, de 17/12/2020, que institui o Protocolo de Investigação para Ilícitos Cibernéticos no âmbito do Poder Judiciário (PGCC/ PJ).

Já no ano de 2021, foram publicadas a Recomendação nº 93, de 06/04/2021, que recomenda o uso da Plataforma de Governança Digital Colaborativa do Poder Judiciário (Connect-Jus); a Recomendação nº 94, de 09/04/2021, que recomenda aos tribunais brasileiros a adoção de medidas incentivadoras da prática de gravação de atos processuais, com vistas à melhoria da prestação jurisdicional; a Recomendação nº 97, de 09/04/2021, que recomenda aos Tribunais de Justiça dos Estados e do Distrito Federal e dos Territórios a utilização de ferramentas tecnológicas para a realização de audiências e atendimentos pelas equipes técnicas, em razão da pandemia mundial por Covid-19, dentre outras recomendações; a Recomendação nº 99, de 21/05/2021, que recomenda aos tribunais e autoridades judiciais a adoção de diretrizes e procedimentos para realização de audiências concentradas para reavaliar as medidas socioeducativas de internação e semiliberdade; a Resolução nº 370, de 28/01/2021 , que Estabelece a Estratégia Nacional de Tecnologia da Informação e Comunicação do Poder Judiciário (ENTIC-JUD); a Resolução nº 371, de 12/02/2021, que Altera a Resolução CNJ nº 227/2016, que regulamenta o teletrabalho no âmbito do Poder Judiciário e dá outras providências; a Resolução nº 372, de 12/02/2021, que regulamenta a criação de plataforma de videoconferência denominada “Balcão Virtual”; a Resolução nº 375, de 02/03/2021, que altera a Resolução CNJ nº 227/2016, que regulamenta o teletrabalho no âmbito do Poder Judiciário e dá outras providências, criando a Equipe de Trabalho Remoto e dando outras providências; a Resolução nº 378, de 09/03/2021, que altera a Resolução CNJ nº 345/2020, que dispõe sobre o “Juízo 100% Digital”; a Resolução nº 383, de 25/03/2021, que cria o Sistema de Inteligência de Segurança Institucional do Poder Judiciário e dá outras providências; a Resolução nº 385, de 06/04/2021, que dispõe sobre a criação dos “Núcleos de Justiça 4.0” e dá outras providências; a Resolução nº 390, de 06/05/2021, que dispõe sobre a extinção de soluções de Tecnologia da Informação e Comunicações e serviços digitais, que foram substituídos ou se encontram inoperantes, fixa regras para a criação de novas soluções de tecnologia e dá outras providências; a Resolução nº 395, de 07/06/2021, que institui a Política de Gestão

da Inovação no âmbito do Poder Judiciário; e a Resolução nº 396, de 07/06/2021, que institui a Estratégia Nacional de Segurança Cibernética do Poder Judiciário (ENSEC-PJ); a Resolução nº 398, de 09/06/2021, que dispõe sobre a atuação dos "Núcleos de Justiça 4.0", disciplinados pela Resolução CNJ nº 385/2021, em apoio às unidades jurisdicionais.

Para os fins desta pesquisa, serão analisadas mais detalhadamente, num momento posterior, a Portaria nº 271 de 04/12/2020, a Resolução nº 332, de 21/08/2020, a Resolução nº 383, de 25/03/2021 e a Resolução nº 395, de 07/06/2021.

Também no ano de 2021 e tendo por base os cinco retromencionados princípios definidos pela OCDE, foi criada a Estratégia Brasileira de Inteligência Artificial (EBIA). A EBIA foi idealizada a partir dos seguintes debates envolvendo a IA:

> [...] a preocupação em estabelecer um ponto de equilíbrio entre: (i) a proteção e a salvaguarda de direitos, inclusive aqueles associados à proteção de dados pessoais e à prevenção de discriminação e viés algorítmico; (ii) a preservação de estruturas adequadas de incentivo ao desenvolvimento de uma tecnologia cujas potencialidades ainda não foram plenamente compreendidas; e (iii) o estabelecimento de parâmetros legais que confiram segurança jurídica quanto à responsabilidade dos diferentes atores que participam da cadeia de valor de sistemas autônomos (EBIA, 2021).

As estratégias propostas pela EBIA na seara jurídica (legislação, regulação e uso ético) são:

> a) Estimular a produção de IA ética financiando projetos de pesquisa que visem aplicar soluções éticas, principalmente nos campos de equidade/não-discriminação (fairness), responsabilidade/prestação de contas (accountability) e transparência (transparency), conhecidas como a matriz FAT. b) Estimular parcerias com corporações que estejam pesquisando soluções comerciais dessas tecnologias de IA ética. c) Estabelecer como requisito técnico em licitações que os proponentes ofereçam soluções compatíveis com a promoção de uma IA ética (por exemplo, estabelecer que soluções de tecnologia de reconhecimento facial adquiridas por órgãos públicos possuam um percentual de falso positivo abaixo de determinado limiar). d) Estabelecer, de maneira multissetorial, espaços para a discussão e definição de princípios éticos a serem observados na pesquisa, no desenvolvimento e no uso da IA. e) Mapear barreiras legais e regulatórias ao desenvolvimento de IA no Brasil e identificar aspectos da legislação brasileira que possam requerer atualização, de modo a promover maior segurança jurídica para o ecossistema digital. f) Estimular ações de transparência e de divulgação responsável quanto ao uso de sistemas de IA, e promover a observância, por tais sistemas, de direitos humanos, de valores democráticos e da diversidade. g) Desenvolver técnicas para identificar e tratar o risco de viés algorítmico. Elaborar política de controle de qualidade de dados para o

> treinamento de sistemas de IA. h) Criar parâmetros sobre a intervenção humana em contextos de IA em que o resultado de uma decisão automatizada implica um alto risco de dano para o indivíduo. i) Incentivar a exploração e o desenvolvimento de mecanismos de revisão apropriados em diferentes contextos de utilização de IA por organizações privadas e por órgãos públicos (EBIA, 2021).

A estratégia busca contribuir para a elaboração de princípios éticos para o desenvolvimento e uso de IA responsáveis, além de promover promover investimentos sustentados em pesquisa e desenvolvimento sobre a temática, remover barreiras à inovação em IA, capacitar e formar profissionais para o ecossistema da IA, estimular a inovação e o desenvolvimento da IA brasileira em ambiente internacional e promover ambiente de cooperação entre os entes públicos e privados, a indústria e os centros de pesquisas para o desenvolvimento da IA. No eixo Legislação, Regulação e Uso Ético, mais especificamente, busca evidenciar a necessidade de se desenvolver parâmetros jurídicos para aplicação da IA (EBIA, 2021).

A Emenda Constitucional 115, de 10 de fevereiro de 2022, altera a Constituição Federal para incluir a proteção de dados pessoais entre os direitos e garantias fundamentais e para fixar a competência privativa da União para legislar sobre proteção e tratamento de dados pessoais (BRASIL, 2022).

Insta destacar também a importância da tendência do cenário legislativo brasileiro.

Nesse ínterim, é possível mencionar o Projeto de Lei do Senado (PLS) 330/2018, que propõe a criação de uma política nacional para a IA no Brasil, estabelecendo princípios, diretrizes e ações governamentais para promover o desenvolvimento e a utilização ética, segura e transparente da IA; o Projeto de Lei da Câmara (PLC) 21/2020, também chamado de Marco Legal da Inteligência Artificial, que busca regulamentar o uso de sistemas de IA na administração pública, estabelecendo critérios para transparência, explicabilidade e responsabilidade no uso dessas tecnologias; o Projeto de Lei (PL) 4649/2020, que propõe a criação de um marco legal para a IA no Brasil, abrangendo aspectos como ética, responsabilidade, privacidade, proteção de dados, transparência e accountability; e o PL 2.338/2023, que dispõe sobre o uso da Inteligência Artificial.

Para além das normativas, existem instrumentos de apoio criados para facilitar o entendimento das diretrizes, como é o caso da Cartilha de Proteção de Dados Pessoais da FIESP, documentos do Confederação Nacional da Indústria (Em busca de soluções: Atributos de autoridade de proteção de dados eficazes e LGPD: o que a sua empresa

precisa saber) e o Guia Como proteger seus dados pessoais, do Ministério da Justiça e Segurança Pública.

2.2.1 Breve explanação acerca do PII na LGPD

Diante da importância conferida pelo próprio ordenamento jurídico brasileiro (à imagem do ordenamento jurídico europeu) à perfilização da pessoa humana e tendo em conta o destaque da informação pessoalmente identificável para o extrativismo de dados, traz-se, a seguir, uma breve discussão acerca do conceito, nos termos da legislação de proteção de dados nacional.

A Lei Geral de Proteção de Dados possui como fundamentos:

> I - o respeito à privacidade;
> II - a autodeterminação informativa;
> III - a liberdade de expressão, de informação, de comunicação e de opinião;
> IV - a inviolabilidade da intimidade, da honra e da imagem;
> V - o desenvolvimento econômico e tecnológico e a inovação;
> VI - a livre iniciativa, a livre concorrência e a defesa do consumidor; e
> VII - os direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais (BRASIL, 2018).

E traz como novidade a classificação dos dados em:

> I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;
> II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;
> III - dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento (BRASIL, 2018).

Sobre o dado anonimizado, a legislação explicita o procedimento, em seu art. 5º, inciso XI, como "anonimização: utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo;" (BRASIL, 2018). O momento do tratamento, nessa seara, deve ser entendido, ainda segundo o mesmo artigo da normativa, como

> X - tratamento: toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração (BRASIL, 2018);

Além disso, essa operação deve ser realizada, segundo o art. 6º da lei, observando-se a boa-fé e os seguintes princípios:

> I - finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;
> II - adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;
> III - necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;
> IV - livre acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;
> V - qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;
> VI - transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;
> VII - segurança: utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;
> VIII - prevenção: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;
> IX - não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;
> 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 (BRASIL, 2018).

Ainda sobre a técnica de anonimização e os dados anonimizados, a lei estabelece:

> Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.
> § 1º A determinação do que seja razoável deve levar em consideração fatores objetivos, tais como custo e tempo necessários para reverter o

> processo de anonimização, de acordo com as tecnologias disponíveis, e a utilização exclusiva de meios próprios.
> § 2º Poderão ser igualmente considerados como dados pessoais, para os fins desta Lei, aqueles utilizados para formação do perfil comportamental de determinada pessoa natural, se identificada.
> § 3º A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.
> Art. 13. Na realização de estudos em saúde pública, os órgãos de pesquisa poderão ter acesso a bases de dados pessoais, que serão tratados exclusivamente dentro do órgão e estritamente para a finalidade de realização de estudos e pesquisas e mantidos em ambiente controlado e seguro, conforme práticas de segurança previstas em regulamento específico e que incluam, sempre que possível, a anonimização ou pseudonimização dos dados, bem como considerem os devidos padrões éticos relacionados a estudos e pesquisas.
> [...]
> Art. 18. O titular dos dados pessoais tem direito a obter do controlador, em relação aos dados do titular por ele tratados, a qualquer momento e mediante requisição:
> IV - anonimização, bloqueio ou eliminação de dados desnecessários, excessivos ou tratados em desconformidade com o disposto nesta Lei; (BRASIL, 2018).

Importante destacar que, para os fins da mencionada lei, considera-se dado pessoal a "informação relacionada a pessoa natural identificada ou identificável", e como dado pessoal sensível o dado pessoal associado à origem racial, étnica, à convicção religiosa, política, à vida sexual, à filiação política, religiosa ou filosófica e o dado genético ou biométrico vinculado à pessoa natural (BRASIL, 2018, recurso online).

Dessa forma, tendo em vista que o dado anonimizado é o dado relativo a titular que não possa ser identificado, seria possível inferir, da leitura dos incisos I, II e III, que uma técnica que se proponha a remover quaisquer traços identificadores de indivíduos de determinada informação deveria se pautar apenas na remoção dos dados intitulados identificadores, englobados pelo inciso I, primeira parte, como é o caso do nome completo, por exemplo.

No entanto, a leitura integral da normativa demonstra que a proposta do legislador é mais extensa: o conceito de anonimização, disposto no inciso XI do mesmo artigo, estabelece que a perda de possibilidade de identificação, ou seja, de associação do dado ao indivíduo, deve ser direta e indireta (BRASIL, 2018, recurso online). O artigo 12 acrescenta, ainda, que o dado anonimizado somente não será considerado dado pessoal quando o processo de anonimização for revertido, utilizando meios exclusivamente próprios, ou quando, com esforços razoáveis, puderem ser revertidos.

Em síntese, a anonimização seria o processo de desidentificação irreversível ou ainda não revertida e que garante a eliminação completa de dados identificados e identificadores do titular. Assim, a técnica incorporaria não apenas a primeira parte do inciso I, do artigo 5º, mas também o dado identificável, contido na segunda parte.

Ocorre que a abordagem probabilística mais recente demonstra que pequenos arranjos de dados (como, por exemplo, apenas 10 URLs) já são suficientes para identificar alguém de forma única, criando "impressões digitais" (HERN, 2017) e sites como o Observatório do Anonimato[30] são capazes de demonstrar isso na prática, apresentando percentuais de reidentificação que aumentam a cada dado colocado pelo usuário.

De posse dessas informações, seria possível traçar inúmeros dados identificáveis de um sujeito. O dado linguístico, por exemplo, apesar de sua ordinariedade e talvez até mesmo por sua natureza usual, fornece padrões tão específicos de utilização dos elementos da linguagem que é capaz de criar marcações de idioleto passíveis de indicação de faixa etária, gênero, orientação sexual, nível de instrução, origem geográfica e outros tantos dados identificáveis.

Por isso, Bioni (2020) conceitua o dado anonimizado como um dado anônimo, incapaz de revelar a identidade de seu titular. E define, ainda, dois modelos de conceituação de dados pessoais: de uma lado, a orientação expansionista, que se define pela delimitação da pessoa identificável, e, de outro, a reducionista, que se define pela delimitação da pessoa identificada.

Essa conceituação é responsável por alargar ou restringir o escopo da aplicação legal, já que a moldura normativa do conceito de pessoa identificada evidencia um vínculo direto, enquanto que o conceito de pessoa identificável, um vínculo indireto. É dizer que, em um modelo expansionista, o dado pode ser definido como qualquer informação relacionada ou relacionável ao seu titular e, num modelo reducionista, o dado somente pode ser a informação diretamente relacionada ao seu titular. Tanto a legislação europeia quanto a brasileira se utilizam do conceito expansionista dos dados.

Nessa toada, a anonimização seria, portanto, "um método cujo mote é gerenciar circunstancialmente a identificabilidade de uma base de dados" (BIONI, 2020). Como cada base de dados é construída com informações específicas e estratégicas direcionadas para determinado objetivo, tal análise deve ser orientada pelas "características de cada dado e a percepção de eles estarem inseridos em uma gama de informações" (BIONI, 2020).

[30] Disponível em https://cpg.doc.ic.ac.uk/observatory/take-the-quiz.

Assim, seria necessário verificar, no processo de anonimização, não somente os dados de identificação direta, mas também os de identificação indireta e, para além disso, o contexto no qual cada dado está inserido, a fim de considerar o grau de identificabilidade da informação quando associada a outra (ou seja, ponderando também sobre a manifesta possibilidade da identificação a partir de metadados).

O Conselho Europeu (2018) apresenta como possíveis técnicas para a eliminação de tais signos identificadores de uma base de dados a supressão, a generalização, a randomização e a pseudoaninimização, mas Bioni (2019) entende que a anonimização é técnica contextual, não havendo uma única forma de parametrizar tal processo.

Considerando que os dados e os metadados se tornaram extremamente valiosos e comercializáveis, compreendemos, como o autor, que a anonimização é um processo contextual voltado para a irreversibilidade. No entanto, o entendimento de que os vínculos de identificação de um titular podem ser completa e irreversivelmente eliminados parece trazer à tona a impossibilidade da efetividade integral da técnica (NARAYANAN, SHMATIKOV, 2010; TEIXEIRA, 2015).

A LGPD traz a conceituação do dado anonimizado, como o "dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento" (BRASIL, 2018, recurso online). O conceito de um dado desprovido dos traços identificadores de seu titular, no entanto, torna-se, semanticamente, controverso e questionável. Se dado é toda informação relativa a um indivíduo que é capaz de identificá-lo (HOUAISS, s/d, recurso online), a definição do dado anonimizado é, *per si*, uma contradição em termos.

Para além, no entanto, da análise filosófica e linguística do termo, a omissão legal acerca da especificidade dos denominados "meios técnicos razoáveis disponíveis" para realizar tal tratamento, torna a temática ainda mais controversa. Isso porque evidências crescentes sugerem que a chamada "desidentificação" do dado não é suficiente para a proteção dos dados pessoais na prática, tendo em vista que poucos excertos de dados pessoais já são o bastante para a identificação de seu titular (ROCHER; MUTHU; DE MONTJOYE, 2021).

Diante do risco inerente de se transmudar em um dado pessoal, (TENE, 2013), o dado anonimizado teria, segundo Bioni (2019), seu espectro cindido daquele do dado pessoal pelo critério da razoabilidade. Dessa forma, não bastaria a mera possibilidade de se atrelar um dado a uma pessoa para a aplicação do termo "identificável" (WP, 2007, p. 1749), mas sim a aplicação de um "esforço razoável" para tanto.

Ainda segundo o autor, a razoabilidade legal possui um eixo objetivo (razoabilidade), composto tanto pelas tecnologias disponíveis quanto pelo custo e o tempo despendidos no esforço realizado para a reversão da anonimização; e um eixo subjetivo (relacionada aos "meios próprios"), composto pela análise individual de capacidade de engenharia reversa do agente de tratamento de dados (BIONI, 2019).

A discussão que orbita a LGPD, portanto, diz respeito à possibilidade da aplicação prática da normativa ante a ausência da previsão de técnicas mais precisas na lei. Essa mesma confrontação será proposta mais adiante, no contexto de uma análise mais geral do quadro jurídico brasileiro.

3 METODOLOGIA

A partir do Relatório de pesquisa: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário – 1ª e 2ª fase (SALOMÃO, 2020), foram levantadas as fontes jurisprudenciais utilizadas na construção dos argumentos aventados pelos documentos.

A escolha pelo relatório se deu por fatores como: elevada estima dos integrantes do Grupo Interinstitucional de Pesquisadores envolvidos no estudo, em especial do ministro do Superior Tribunal de Justiça e corregedor nacional de Justiça Luis Felipe Salomão; reconhecimento da seriedade Fundação Getúlio Vargas como fundação, particularmente no meio jurídico; completude da pesquisa, que envolve o Conselho Nacional de Justiça, o Supremo Tribunal Federal, o Superior Tribunal de Justiça, o Tribunal Superior do Trabalho, os Tribunais Regionais do Trabalho, os Tribunais Regionais Federais e os Tribunais de Justiça; e facilidade de acesso à documentação, principalmente no que se refere à gratuidade do compilado.

Foram levantadas as seguintes fontes: a GDPR PARLAMENTO EUROPEU, 2016), a Resolução 2015/2103 (INL) (PARLAMENTO EUROPEU, 2015), A Comunicação "Inteligência artificial para a Europa (COMISSÃO EUROPEIA, 2018), o Plano Coordenado para a Inteligência Artificial (COMISSÃO EUROPEIA, 2021), *Building Trust in Human-Centric Artificial Intelligence* (COMISSÃO EUROPEIA, 2019), Orientações Éticas para uma IA de Confiança (AI HLEG, 2019), Recomendações sobre a Ética da IA (UNESCO, 2021), Aumentar a confiança numa inteligência artificial centrada no ser humano (COMISSÃO EUROPEIA, 2019), Declaração sobre Ética e Proteção de Dados em Inteligência Artificial (ICDPPC, 2018), Carta Ética Europeia sobre o uso de IA nos sistemas judiciais (COMISSÃO EUROPEIA, 2018), Digitalização para uma justiça melhor (CEPEJ, 2021), Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica (PARLAMENTO EUROPEU, 2019), Livro Branco sobre a inteligência artificial (COMISSÃO EUROPEIA, 2020), Proposta de Regulamento sobre Inteligência Artificial (COMISSÃO EUROPEIA, 2022), princípios da OCDE (OCDE, 2019), Marco Civil (BRASIL, 2014), Portaria nº 271 de 04/12/2020 (CNJ, 2020), Resolução nº 332 do CNJ (CNJ, 2020), Resolução nº 395 do CNJ (CNJ, 2021), PL 21/2020 (BRASIL, 2020), PL 2.338 (BRASIL, 2023).

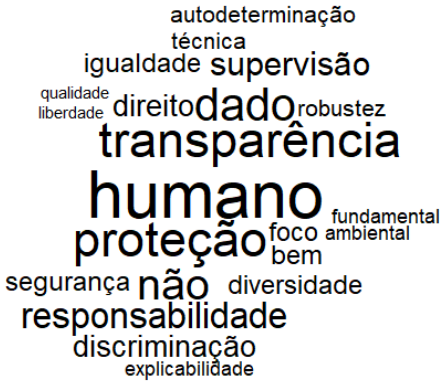
De tal exposição, foi confeccionado um excel com duas planilhas: legislação nacional e legislação internacional. Dentro de cada planilha, foram feitas colunas contendo as legislações analisadas e, nas células, os princípios previstos em cada legislação.

A partir da análise linguística dos documentos no programa Iramuteq, foram traçados 40 princípios preponderantes no âmbito internacional: lealdade, equidade, transparência, explicitude, legitimidade, proteção de dados, liberdade, autodeterminação, responsabilidade, supervisão humana, IA responsável, não perfilização, foco no humano, IA confiável, não discriminação, sustentabilidade, dignidade humana, democracia, Estado de Direito, robustez técnica, diversidade, não discriminação, Bem-estar social, Bem-estar ambiental, legalidade, não maleficência, igualdade, explicabilidade, respeito, direitos humanos, sociedades pacíficas, proporcionalidade, segurança, transparência, conscientização, alfabetização, governança, IA de confiança, direitos fundamentais, imparcialidade; e 19 no âmbito nacional: liberdade, proteção de dados, imparcialidade, estabilidade, responsabilidade, economicidade, interoperabilidade tecnológica, transparência, informação, direitos fundamentais, segurança jurídica, igualdade, não discriminação, diversidade, inovação, foco no humano, acessibilidade, desburocratização, Bem-estar ambiental.

Restaram 57 valores únicos: lealdade, equidade, transparência, explicitude, legitimidade, proteção de dados, liberdade, autodeterminação, responsabilidade, supervisão humana, IA responsável, não perfilização, foco no humano, IA confiável, não discriminação, sustentabilidade, dignidade humana, democracia, Estado de Direito, robustez técnica, diversidade, Bem-estar social, Bem-estar ambiental, legalidade, não maleficência, igualdade, explicabilidade, respeito, direitos humanos, sociedades pacíficas, proporcionalidade, segurança, conscientização, alfabetização, governança, IA de confiança, direitos fundamentais, imparcialidade, liberdade, proteção de dados, imparcialidade, estabilidade, responsabilidade, economicidade, interoperabilidade tecnológica, transparência, informação, direitos fundamentais, segurança jurídica, igualdade, não discriminação, diversidade, inovação, foco no humano, acessibilidade, desburocratização, Bem-estar ambiental.

O mencionado Excel foi passado a um arquivo no formato de base de dados do software Iramuteq, tornando-se fonte apropriada para análise de *corpus*.

O Iramuteq gerou a seguinte nuvem de palavras:



Fonte: Iramuteq, 2023.

A análise do *corpus* demonstrou que os princípios que mais aparecem na seara nacional são: responsabilidade, proteção de dados e transparência; já na seara internacional, são: proteção de dados, transparência e supervisão humana; na análise conjunta, os princípios que mais apareceram foram: proteção de dados, transparência, responsabilidade, supervisão humana e não discriminação.

Em seguida, realizou-se uma análise qualitativa a fim de verificar se os princípios proteção de dados, transparência, responsabilidade, supervisão humana e não discriminação possuem o mesmo conteúdo e/ou interpretação em todos os documentos em que são mencionados e se, no cenário nacional, as normativas específicas e a atuação jurídica são coerentes com a dicção a que se propõem os princípios no Brasil.

4 RECORTE DOS PRINCÍPIOS DO ARCABOUÇO LEGAL CONSTITUÍDO

Diante do estado da arte até aqui delineado, o presente estudo se dedica, nesta seção, a se debruçar sobre os documentos selecionados a partir do recorte de seu objeto, a fim de apresentar uma análise sintetizada do arcabouço legal sob a ótica das recomendações e diretrizes existentes na literatura específica.

O objetivo será traçar tendências legislativas e observar as principais inferências possíveis de serem feitas a partir da análise dos princípios dos documentos mais relevantes para este estudo. O recorte pautado em princípios tem relação com a generalização desse tipo de norma, considerada mandamento de otimização (ALEXY, 2005).

Da seção de princípios da GDPR, a leitura do dispositivo permite destacar a preocupação com o tratamento de dados pessoas (que deve se dar de forma lícita, leal e transparente, seguindo a lealdade, equidade e transparência) e a coleta pautada na explicitude e legitimidade dos fins; a utilização deve ser adequada aos fins necessários e a conservação deve permitir a identificação não mais do que o necessários para as finalidades do tratamento da informação. A GDPR apresenta também seção dedicada ao consentimento do titular e ao processamento de dados pessoais (proibindo a revelação de origem racial ou étnica, as opiniões políticas, as convicções religiosas ou filosóficas ou a filiação sindical, bem como o tratamento de dados genéticos, dados biométricos para fins de identificação unívoca de uma pessoa singular, dados relativos à saúde ou dados relativos à vida sexual ou orientação sexual de uma pessoa singular) (PARLAMENTO EUROPEU, 2016).

Já a Resolução 2015/2103 (INL) apresenta as seguintes características de uma IA: aquisição de autonomia através de sensores e/ou da troca de dados com o seu ambiente (interconetividade) e da troca e análise desses dados; autoaprendizagem com a experiência e com a interação (critério opcional); um suporte físico mínimo; adaptação do seu comportamento e das suas ações ao ambiente; inexistência de vida no sentido biológico do termo. Além disso, apresenta os seguintes princípios gerais relativos ao desenvolvimento da robótica e da inteligência artificial para utilização civil: análise da segurança, da saúde e da proteção humanas; da liberdade, da privacidade, da integridade e da dignidade; da autodeterminação e da não discriminação, e da proteção dos dados pessoais; atualização e complementação do quadro jurídico de acordo com a complexidade robótica vigente; presença de um quadro ético orientador, claro, rigoroso e eficiente pautado nos princípios de beneficência, não-maleficência, autonomia e justiça, nos princípios e valores

consagrados no artigo 2.º do Tratado da União Europeia e na Carta dos Direitos Fundamentais, tais como a dignidade do ser humano, a igualdade, a justiça e a equidade, a não discriminação, o consentimento esclarecido, o respeito pela vida privada e familiar e a proteção de dados, bem como em outros princípios e valores subjacentes do direito da União, como a não estigmatização, a transparência, a autonomia, a responsabilidade individual e a responsabilidade social, e em códigos e práticas éticas existentes. Por fim, o documento realça o princípio da transparência e a ameaça à privacidade (PARLAMENTO EUROPEU, 2015).

A Comunicação "Inteligência artificial para a Europa"[31] traz uma proposta sustentável de quadro ético e jurídico para os valores nos quais as novas tecnologias se baseiam e, considerando as novas questões advindas das aplicações da IA, defende a manutenção dos direitos fundamentais e de princípios éticos como a responsabilização e a transparência. O documento enfatiza a aplicação da GDPR no que se refere à proteção de dados pessoais e disposições sobre tomada de decisões com base unicamente no tratamento automatizado, incluindo, ainda, a definição de perfis e afirma que o princípio orientador da IA deverá ser sempre o do desenvolvimento da IA responsável, colocando o homem no centro do processo, além do princípio da inovação (um conjunto de ferramentas e orientações desenvolvido para garantir que todas as iniciativas da Comissão são favoráveis à inovação) (COMISSÃO EUROPEIA, 2018).

No que se refere ao "Plano Coordenado para a Inteligência Artificial", seu objetivo está centrado no desenvolvimento de uma estrutura de políticas para garantir a confiança nos sistemas de IA, com uma abordagem centrada no ser humano e na proteção dos valores da UE e nos direitos fundamentais, como não discriminação, privacidade e proteção de dados, e o uso sustentável e eficiente de recursos estão entre os princípios-chave que orientam a abordagem europeia. Além disso, o documento enfatiza a conformidade com a legislação de proteção de dados e com os princípios éticos e regras de competição para identificar novos conhecimentos e apoiar pesquisa e tomada de decisões na iniciativa, mencionando as Orientações Éticas para uma IA de Confiança e a Lista de Avaliação para IA Confiável (ALTAI), produzidas pelo Grupo de Peritos de Alto Nível em IA (AI HLEG) (COMISSÃO EUROPEIA, 2021).

O documento *Building Trust in Human-Centric Artificial Intelligence* se baseia nos valores da dignidade humana, liberdade, democracia, equidade, Estado de Direito e

[31] Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52018DC0237. Acesso em 23 de jul. de 2023.

respeito aos direitos humanos, pautando-se em sete princípios: Agência humana e supervisão, Robustez técnica e segurança, Privacidade e governança de dados, Transparência, Diversidade, não discriminação e justiça, Bem-estar social e ambiental e Responsabilidade (COMISSÃO EUROPEIA, 2019).

As Orientações Éticas para uma IA de Confiança, por sua vez, pauta os três componentes de uma IA de confiança na legalidade, na ética e na solidez técnica; quanto ao critério legal, este deve respeitar a autonomia humana, se basear na prevenção de danos, na equidade e na explicabilidade (AI HLEG, 2019).

As Recomendações sobre a Ética da IA, publicadas pela UNESCO, trazem como valores: respeito, proteção e promoção dos direitos humanos, das liberdades fundamentais e da dignidade humana (prosperidade ambiental e ecossistêmica, garantir diversidade e inclusão, viver em sociedades pacíficas, justas e interconectadas); e como princípios: proporcionalidade e não causar dano, segurança e proteção, justiça e não discriminação, sustentabilidade, direito à privacidade e proteção de dados, transparência e explicabilidade, supervisão humana e determinação, conscientização e alfabetização, responsabilidade e prestação de contas, governança e colaboração adaptáveis e com múltiplas partes interessadas (UNESCO, 2021).

O documento "Aumentar a confiança numa inteligência artificial centrada no ser humano"[32] traz a ideia de uma IA de confiança, coerente com a legislação, com a diversidade e os direitos fundamentais e que tem por objetivo capacitar pessoas e não substituí-las. Os requisitos para o desenvolvimento de tal modelo, segundo o documento, seriam: Iniciativa e controlo por humanos, Robustez e segurança, Privacidade e governação dos dados, Transparência, Diversidade, não discriminação e equidade, Bem-estar societal e ambiental e Responsabilização; e, para alcançá-los, deveriam ser garantidos os direitos à privacidade, à determinação humana, à identificação, à justiça, à *accountability*, à acurácia, à qualidade de dados, à segurança pública, à cibersegurança, à proibição ao perfilamento secreto, à proibição ao scoring unitário e à obrigação rescisão (COMISSÃO EUROPEIA, 2019).

A Declaração sobre Ética e Proteção de Dados em Inteligência Artificial tem como princípios basilares: optar pelo uso do marco internacional dos Direitos Humanos para avaliar os efeitos da IA; reconhecer explicitamente grupos em condição especial de vulnerabilidade em decorrência dos sistemas de IA; explicitar responsabilidades dos

[32] Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52019DC0168. Acesso em 23 de jul. de 2023.

Estados; reconhecer as tensões que a IA introduz no sistema tradicional de proteção de dados e, desde esse ponto, avançar em soluções acordadas; explicitar o trabalho das autoridades que tenham como função supervisionar sistemas de IA; lidar com as forças oligopólicas do mercado e seu efeito na IA e desenvolver a segurança dos sistemas de IA e de seus resultados (ICDPPC, 2018).

A Carta Ética Europeia sobre o uso de IA nos sistemas judiciais tem por princípios: o respeito aos direitos fundamentais, a não-discriminação, a qualidade e segurança, a transparência, imparcialidade e equidade e o controle do usuário (COMISSÃO EUROPEIA, 2018).

O documento "Digitalização para uma justiça melhor", traz as orientações de eficiência, transparência, justiça colaborativa, justiça humana, justiça centrada nas pessoas, justiça informada e CEPEJ responsável e retroativo, de modo que garanta a visibilidade das ferramentas desenvolvidas (CEPEJ, 2021).

O Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica traz os seguintes aspectos éticos: tecnologia centrada no ser humano; valores incorporados na tecnologia – ética desde a conceção; o processo de decisão – limites à autonomia da inteligência artificial e da robótica; transparência, enviesamento e explicabilidade dos algoritmos (PARLAMENTO EUROPEU, 2019).

O Livro Branco sobre a inteligência artificial reafirma os requisitos essenciais de uma IA de confiança, pautado nas orientações do Grupo de Peritos de Alto Nível e a necessidade de um quadro regulamentar europeu claro e sólido que reforce os princípios e valores da UE a fim de proteger os cidadãos e criar um mercado interno sem atritos. Para tanto, apresenta como questões para as aplicações de IA de alto risco: dados de treino; conservação de registos e de dados; prestação de informações; robustez e exatidão; supervisão humana e requisitos específicos para determinadas aplicações de IA, tais como as utilizadas para fins de identificação biométrica à distância (COMISSÃO EUROPEIA, 2020).

A Proposta de Regulamento sobre Inteligência Artificial cita a EBIA, apresentando soluções como a capAI, uma ferramenta desenvolvida na Califórnia que fornece às organizações orientações práticas sobre como traduzir princípios éticos de alto nível em critérios verificáveis que ajudam a moldar o design, desenvolvimento, implantação e uso da IA ética; além de fornecer informações sobre como seus dados são usados e não devem estar sujeitos a práticas abusivas, de forma os usuários devem saber por que e como um

sistema de IA fez sua determinação, as pessoas devem ter a opção de optar por não tomar decisões de IA e recorrer a um ser humano se o sistema apresentar um erro, falhar ou se elas quiserem contestar a decisão. A *Australia's Ethics Framework*, outra estrutura apresentada, inclui princípios de ética da IA para ajudar a reduzir o risco de impactos negativos da IA e garantir que o uso da IA seja sustentado por padrões de boa governança (são eles: 1- Não prejudicar; 2- Conformidade Regulatória; 3- Proteção e privacidade; 4- Equidade; 5- Transparência e Explicabilidade; 6- Contestação; e 7- Responsabilidade). O *Draft AI R&D guidelines for international discussion*, outro documento citado, no qual o Japão reconhecia os grandes benefícios da IA e o *Utilization Guidelines*, fornece orientações para usuários de IA (quais sejam: 1) Princípio da colaboração; 2) Princípio da transparência; 3) Princípio da controlabilidade; 4) Princípio da segurança; além dos princípios sociais: 1- princípio centrado no ser humano; 2- princípio da educação e alfabetização; 3- princípio da proteção da privacidade; 4- princípio de garantia da segurança; 5- princípio da concorrência leal; 6- princípios de justiça, responsabilidade e transparência; e 7- princípio da inovação) (COMISSÃO EUROPEIA, 2022).

A OCDE, como mencionado anteriormente, possui cinco princípios orientadores: 1. A IA deve beneficiar não só as pessoas, mas também o nosso planeta, contemplando o crescimento inclusivo, o desenvolvimento sustentável e o bem-estar geral; 2. Os sistemas de IA devem ser concebidos de forma a respeitar o estado de direito, os direitos humanos, os valores democráticos e a diversidade. Devem igualmente incluir salvaguardas adequadas - por exemplo, permitir a intervenção humana, quando necessário - para garantir uma sociedade justa e equitativa; 3. Os sistemas de IA devem ser transparentes e assegurar uma divulgação responsável, de modo a garantir que as pessoas compreendam os resultados provenientes do seu uso e possam geri-los corretamente; 4. Os sistemas de IA devem funcionar de forma robusta e segura, durante todo o seu ciclo de vida. Os seus potenciais riscos devem ser continuamente avaliados e geridos; e 5. As organizações e os indivíduos que desenvolvem, implantam ou operam sistemas de IA devem assegurar o seu bom funcionamento, em conformidade com os princípios acima referidos (OCDE, 2019).

No cenário nacional, o marco civil traz os seguintes princípios: garantia da liberdade de expressão, comunicação e manifestação de pensamento; proteção da privacidade e dos dados pessoais; preservação da neutralidade de rede, além da estabilidade, segurança e funcionalidade da rede; responsabilização; preservação da natureza participativa da rede e liberdade dos modelos de negócios, desde que não conflitem com os demais princípios. O art. 7º da legislação apresenta também direitos de privacidade e inviolabilidade, além do

tratamento de dados, consentimento, direito à publicidade e acessibilidade (BRASIL, 2014).

A Portaria nº 271 de 04/12/2020 afirma que são considerados como de inteligência artificial os projetos voltados a:

> I – criar soluções para automação dos processos judiciais e administrativos e de rotinas de trabalho da atividade judiciária;
> II – apresentar análise da massa de dados existentes no âmbito do Poder Judiciário; e
> III – prover soluções de apoio à decisão dos magistrados ou à elaboração de minutas de atos judiciais em geral (CNJ, 2020).

Afirmando que a pesquisa e desenvolvimento em matéria de inteligência artificial observará a economicidade, a promoção da interoperabilidade tecnológica, a adoção de tecnologias, padrões e formatos abertos e livres, o acesso à informação, a transparência, a capacitação humana, a celeridade processual e o estabelecimento de mecanismos de governança colaborativa e democrática. A normativa também enfatiza a disponibilização dos dados utilizados para treinamento no modelo; a responsabilidade sobre a preservação, no momento da disponibilização, da anonimização; e a necessidade de rastreamento e auditoria das predições dos modelos inteligentes (CNJ, 2020).

A Resolução nº 332 do CNJ enfatiza a necessidade de compatibilizar a IA aos direitos fundamentais, à segurança jurídica e à igualdade de tratamento aos casos absolutamente iguais. Além disso, defende a utilização de amostras representativas no treinamento de modelos e as cautelas quanto aos dados sensíveis, trazendo princípios como a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo, com criação de condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos. A resolução traz a necessidade de transparência, incluindo indicação dos objetivos, resultados e riscos do modelo; e a necessidade de observância a regras de governança, principalmente em modelos adotados pelo Poder Judiciário:

> Art. 25. Qualquer solução computacional do Poder Judiciário que utilizar modelos de Inteligência Artificial deverá assegurar total transparência na prestação de contas, com o fim de garantir o impacto positivo para os usuários finais e para a sociedade.
> Parágrafo único. A prestação de contas compreenderá:
> I – os nomes dos responsáveis pela execução das ações e pela prestação de contas;

> II – os custos envolvidos na pesquisa, desenvolvimento, implantação, comunicação e treinamento;
> III – a existência de ações de colaboração e cooperação entre os agentes do setor público ou desses com a iniciativa privada ou a sociedade civil;
> IV – os resultados pretendidos e os que foram efetivamente alcançados;
> V – a demonstração de efetiva publicidade quanto à natureza do serviço oferecido, técnicas utilizadas, desempenho do sistema e riscos de erros (CNJ, 2020).

A Resolução nº 395 do CNJ traz como princípios da gestão de inovação no Poder Judiciário: a cultura da inovação; o foco no usuário; a participação; a colaboração; o desenvolvimento humano; a acessibilidade; a sustentabilidade socioambiental; o desenvolvimento sustentável; a desburocratização; a transparência (CNJ, 2021).

O PL 21/2020 traz alguns princípios não mencionados em outros documentos, como o aumento da competitividade e da produtividade brasileira; a inserção competitiva do Brasil nas cadeias globais de valor; o reconhecimento da natureza digital, transversal e dinâmica da IA; o estímulo à autorregulação, mediante adoção de códigos de conduta e de guias de boas práticas; a defesa nacional. Especificamente sobre a aplicação da IA, menciona princípios já apresentados em outras normativas, como a finalidade benéfica, a centralidade do ser humano, a não discriminação, a busca pela neutralidade, a transparência, a segurança e prevenção, a inovação responsável, a disponibilidade de dados, a participação social e interdisciplinar, a análise de impacto regulatório e a responsabilidade (BRASIL, 2020).

O PL 2.338, por sua vez, traz como princípios da IA:

> Art. 2º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial no Brasil têm como fundamentos:
> I – a centralidade da pessoa humana;
> II – o respeito aos direitos humanos e aos valores democráticos;
> III – o livre desenvolvimento da personalidade;
> IV – a proteção ao meio ambiente e o desenvolvimento sustentável;
> V – a igualdade, a não discriminação, a pluralidade e o respeito aos direitos trabalhistas;
> VI – o desenvolvimento tecnológico e a inovação;
> VII – a livre iniciativa, a livre concorrência e a defesa do consumidor;
> VIII – a privacidade, a proteção de dados e a autodeterminação informativa;
> IX – a promoção da pesquisa e do desenvolvimento com a finalidade de estimular a inovação nos setores produtivos e no poder público; e
> X – o acesso à informação e à educação, e a conscientização sobre os sistemas de inteligência artificial e suas aplicações.
> Art. 3º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios:
> I – crescimento inclusivo, desenvolvimento sustentável e bemestar;

> II – autodeterminação e liberdade de decisão e de escolha;
> III – participação humana no ciclo da inteligência artificial e supervisão humana efetiva;
> IV – não discriminação;
> V – justiça, equidade e inclusão;
> VI – transparência, explicabilidade, inteligibilidade e auditabilidade;
> VII – confiabilidade e robustez dos sistemas de inteligência artificial e segurança da informação;
> 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;
> XI – prevenção, precaução e mitigação de riscos sistêmicos derivados de usos intencionais ou não intencionais e de 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 (BRASIL, 2023).

Na análise conjunta, os princípios que mais apareceram foram: proteção de dados, transparência, responsabilidade, supervisão humana e não discriminação. Tendo em conta os princípios mais citados pela literatura específica, cabe questionar se estes possuem o mesmo conteúdo ou a mesma interpretação em todos os documentos em que são mencionados.

5 ANÁLISE QUALITATIVA DO RECORTE CONSTITUÍDO

Tendo em conta os princípios mais citados pela literatura específica, cabe questionar se estes possuem o mesmo conteúdo ou a mesma interpretação em todas os documentos em que são mencionados e se, no cenário nacional, as normativas específicas são coerentes com a dicção a que se propõem os princípios no Brasil.

5.1 DO PRINCÍPIO DA RESPONSABILIDADE

O princípio da responsabilidade, na GDPR, é explicitamente citado no art. 5º da legislação, que afirma que "o responsável pelo tratamento é responsável pelo cumprimento do disposto no n.o 1 [princípios de tratamento de dados] e tem de poder comprová-lo («responsabilidade»)" (GDPR, 2016).

A Resolução 2015/2103 (INL), ao tratar da responsabilidade, aborda a responsabilidade civil como questão crucial diante do paradigma da mudança do uso das IA's, aventando, ainda, a insuficiência do arcabouço jurídico contemporâneo em relação à temática. Na seção de princípios éticos, destaca que a responsabilidade deve ser compreendida sob o viés individual e social. Na seção dedicada ao tema, o documento afirma que a questão tem de ser resolvida ao nível da União, "a fim garantir o mesmo nível de eficácia, de transparência e de coerência na execução da segurança jurídica em toda a União", considerando, ainda, que

> [...] em princípio, uma vez identificadas as partes às quais cabe, em última instância, a responsabilidade, esta deve ser proporcionada em relação ao nível efetivo de instruções dadas ao robô e ao nível da sua autonomia, de modo a que quanto maior for a capacidade de aprendizagem ou de autonomia de um robô, e quanto mais longa for a «educação» do robô, maior deve ser a responsabilidade do «professor»; observa, em especial, que as competências resultantes da «formação» dada a um robô não devem ser confundidas com as competências estritamente dependentes das suas capacidades de autoaprendizagem, quando se procura identificar a pessoa à qual se atribui efetivamente o comportamento danoso do robô; observa que, pelo menos na fase atual, a responsabilidade deve ser imputada a um ser humano, e não a um robô; (PARLAMENTO EUROPEU, 2015)

O documento afirma, ainda, que a abordagem utilizada nos danos causados por agente não humano, deve ser a da responsabilidade objetiva ou a de gestão de riscos, de forma que não se limite o tipo, a extensão ou as formas de compensação a se indenizar.

Afirma, ainda, que as decisões políticas sobre tais regras de responsabilidade deverão ser tomadas com base em informações de um projeto dedicado à robótica e à neurociência, com cientistas e especialistas da área (PARLAMENTO EUROPEU, 2015).

A Comunicação “Inteligência artificial para a Europa”, quando cita a responsabilidade, traz preocupação quanto à garantia de um quadro jurídico apropriado para a temática, garantindo a confiança em torno da utilização da IA, se propondo a publicar um relatório sobre as implicações mais abrangentes em termos de matéria de segurança e responsabilidade decorrentes da IA, da IoT e da robótica (COMISSÃO EUROPEIA, 2018).

O Plano Coordenado para a Inteligência Artificial tem como um dos objetivos propor adaptações legislativas no quadro jurídico a respeito do tema. Traz também a aproximação da responsabilidade à confiança dos sistemas robóticos e afirma que o quadro de responsabilidade deve "assegurar que as vítimas que sofrem danos na sua vida, saúde ou património em resultado da utilização de novas tecnologias têm acesso às mesmas indemnizações que as vítimas de outras tecnologias" (COMISSÃO EUROPEIA, 2021).

O documento *Building Trust in Human-Centric Artificial Intelligence* afirma que a responsabilidade e a responsabilização devem ser asseguradas, antes e depois da aplicação dos sistemas de IA. A responsabilidade, nesta carta, se aproxima da auditabilidade dos sistemas (relacionada aos relatórios de avaliação de impacto, que contribuem para a fiabilidade da tecnologia e que devem ser proporcionais à dimensão dos riscos representados por cada modelo inteligente) e à reparação adequada quando da ocorrência de impactos "adversos e injustos" (COMISSÃO EUROPEIA, 2019). Esta carta lança bases para o documento "Aumentar a confiança numa inteligência artificial centrada no ser humano", do mesmo ano, que possui a mesma inteligência normativa.

No mesmo sentido, as Orientações Éticas para uma IA de Confiança afirmam que o requisito de responsabilização está estreitamente relacionado com o princípio da equidade, já que exige que "sejam criados mecanismos para garantir a responsabilidade e a responsabilização pelos sistemas de IA e os seus resultados", relacionando o princípio também à auditabilidade (AI HLEG, 2019).

O documento As Recomendações sobre a Ética da IA prevê que os Estados-membros devem garantir que sempre seja possível atribuir responsabilidade aos modelos de IA, categorizando-a em responsabilidade legal (ou jurídica) e responsabilidade ética. Nesta carta, a responsabilidade aparece associada, além de à auditabilidade, ao controle e ao princípio da supervisão humana, que é classificado em privado e público:

> É possível que, às vezes, as pessoas decidam confiar em sistemas de IA por motivos de eficácia, mas a decisão de ceder o controle em contextos limitados continua sendo de seres humanos, pois estes podem recorrer àqueles sistemas para tomar decisões e agir, mas um sistema de IA jamais poderá substituir a responsabilidade e a prestação de contas finais humanas. Como regra, decisões de vida e morte não devem ser transferidas a sistemas de IA (UNESCO, 2021).

Nos documentos A Declaração sobre Ética e Proteção de Dados em Inteligência Artificial, A Carta Ética Europeia sobre o uso de IA nos sistemas judiciais e Digitalização para uma justiça melhor, o princípio aparece poucas vezes e as menções não apresentam correlação com o objeto debatido neste estudo.

A citação mais enfática do Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica sobre o princípio da responsabilidade é a de que "Considerando que a utilização da inteligência artificial, especialmente no setor da saúde, se deve basear sempre no princípio de responsabilidade de que «é o Homem que comanda a máquina». A responsabilidade aparece também integrada ao modelo ético de concepção da IA, aparentando sua definição estar intrinsecamente relacionada ao controle humano (PARLAMENTO EUROPEU, 2019).

No Livro Branco sobre a IA, a Comissão Europeia se utiliza dos termos da Diretiva Responsabilidade pelos Produtos para atribuir responsabilidade ao fabricante da IA, enumerando exceções e introduzindo a possibilidade de inserção de novas disposições que abranjam explicitamente eventuais novos riscos apresentados pelas tecnologias digitais emergentes (COMISSÃO EUROPEIA, 2020).

Na Proposta de Regulamento sobre Inteligência Artificial, o princípio da responsabilidade aparece também relacionado à supervisão e ao controle humano, havendo ênfase do documento à possibilidade de responsabilização dos utilizadores dos sistemas de IA quando se tratar de modelos de alto risco (COMISSÃO EUROPEIA, 2022).

Na *Recommendation of the Council on OECD Legal Instruments Artificial Intelligence*, o princípio da responsabilidade é tido como complementar aos demais princípios, ao crescimento inclusivo, desenvolvimento sustentável e bem-estar; valores e justiça centrados no ser humano; transparência e explicabilidade; robustez, segurança e proteção da IA é equiparado à *accountability* (OCDE, 2019).

No Marco Civil da Internet e nas Resoluções nº 332 e nº 395 do CNJ o princípio não é explicitamente mencionado.

Na Portaria nº 271 de 04/12/2020, a responsabilidade aparece explicitamente no art. 5º, quando o documento assegura que "a administração da plataforma de inteligência artificial do Poder Judiciário, abrangendo seus subsistemas e modelos, ficará sob a responsabilidade e coordenação do CN". Além disso, a responsabilidade sobre a proteção de dados pessoais aparece como um ônus do órgão criador e/ou mantenedor de cada modelo de IA (CNJ, 2020).

No PL 21/2020, a responsabilidade por proteger os modelos de IA contra ameaças de segurança cibernética é atribuída aos agentes de desenvolvimento e de operação de tais sistemas (BRASIL, 2020).

No PL 2.338, a atribuição da responsabilidade aparece no desenvolvimento, implementação e uso dos sistemas de IA: o princípio é associado à rastreabilidade das decisões dos modelos, que é equiparada a um procedimento de prestação de contas; por outro lado, a responsabilidade é interpretada também como reparação (responsabilidade civil), estando o fornecedor ou operador de sistema de inteligência artificial obrigado a reparar o dano (patrimonial, moral, individual ou coletivo) integralmente, independentemente do grau de autonomia do modelo, sendo aplicável, no que couber, a legislação consumerista (BRASIL, 2023)

5.2 DO PRINCÍPIO DA PROTEÇÃO DE DADOS

O princípio aparece, na Resolução 2015/2103 (INL), na abordagem de dispositivos que se intercomunicam com base de dados sem intervenção de humanos e na solicitação de clarificação de normas específicas para o uso de câmeras e sensores. Além disso, é relacionado aos princípios da minimização dos dados, da limitação da finalidade e de mecanismos de controle da transparência para titulares de dados como bases para a proteção de dados, tendo em conta também a livre circulação de dados como fundamento para a economia digital (PARLAMENTO EUROPEU, 2015).

A Comunicação “Inteligência artificial para a Europa” defende uma maior disponibilidade de dados de bases privadas, devendo haver proteção dos dados pessoais e reutilização de dados não pessoais (inclusive para treino de sistemas de IA) (COMISSÃO EUROPEIA, 2018).

O “Plano Coordenado para a Inteligência Artificial” relaciona o princípio a "questões relativas à governação, segurança, proteção de dados e privacidade, qualidade, infraestruturas e interoperabilidade de dados, saúde digital e IA", a fim também de assegurar

o fluxo livre e seguro de dados e promover a adoção de soluções digitais e da IA (COMISSÃO EUROPEIA, 2021).

O documento *Building Trust in Human-Centric Artificial Intelligence* cita os conceitos de proteção de dados *by default* (por padrão) e *by design* (desde o desenho ou a concepção) para afirmar que a privacidade e a proteção de dados (tratados como definições equiparadas) devem ser garantidas em todas as fases do ciclo de vida dos modelos de IA. O princípio também é tratado em conjunto à qualidade dos sistemas de IA, para garantir a não discriminação pela criação de vieses, e à regulamentação e controle do acesso aos dados (COMISSÃO EUROPEIA, 2019).

As Orientações Éticas para uma IA de Confiança também tratam a privacidade e a proteção dos dados conjuntamente, abordando-as em constante relação com a informação dada ao usuário e inicialmente fornecida pelo utilizador (AI HLEG, 2019).

As Recomendações sobre a Ética da IA afirmam que a privacidade é um direito essencial para proteger a dignidade, a autonomia e a capacidade de ação humanas. A proteção dos dados é intrinsecamente relacionada ao exercício dos direitos pelos titulares dos dados, a fim de garantir um objetivo legítimo e uma base jurídica válida para o processamento dos dados pessoais, incluindo o consentimento consciente (UNESCO, 2021).

A Declaração sobre Ética e Proteção de Dados em Inteligência Artificial diz que a escassa ou inexistente transparência acerca do contexto em se coletaram os dados que alimentam a tomada de decisões das máquinas tornam-nas inescrutáveis e que, de outro lado, o problema da autorização explícita dos titulares de dados para utilizações secundárias (dificuldade do consentimento informado) gera tensões no sistema tradicional de proteção de dados. Dessa forma, o princípio da proteção de dados pessoais se associa à transparência (ICDPPC, 2018).

A Carta Ética Europeia sobre o uso de IA nos sistemas judiciais critica a possibilidade de um modelo econômico de dados de jurisprudência pública ser tratado gratuitamente pelos setores privados; além disso, relaciona o princípio da proteção de dados pessoais diretamente ao princípio da precaução para avaliação de eventuais riscos (COMISSÃO EUROPEIA, 2018).

Não há menção enfática sobre o princípio no documento Digitalização para uma justiça melhor (CEPEJ, 2021).

O Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica cita a necessidade de medidas de minimização da discriminação e parcialidade algorítmicas, estando a proteção de dados ligada ao

tratamento transparente de dados pessoais; há também menção à proteção de dados por definição e desde a concepção, juntamente à limitação da finalidade, da armazenagem, à exatidão e à minimização de dados (PARLAMENTO EUROPEU, 2019).

O Livro Branco sobre Inteligência Artificial afirma que a utilização da IA oferece riscos aos valores da União (direitos à liberdade de expressão, à liberdade de reunião, à dignidade humana, à não discriminação em virtude do sexo, origem racial ou étnica, religião ou crença, deficiência, idade ou orientação sexual, tal como aplicável em determinados domínios, proteção de dados pessoais e da vida privada, etc) e, mais do que aos dados pessoais que podem desaguar na identificação inequívoca de uma pessoa singular, mas também a conjuntos de dados que, por si só, não incluem dados pessoais (COMISSÃO EUROPEIA, 2020).

A menção mais relevante do princípio na Proposta de Regulamento sobre Inteligência Artificial é a de que a proteção de dados pessoais e a privacidade, aliados a outros valores da União (como dignidade do ser humano, liberdade, igualdade, democracia, não discriminação, etc) devem lançar bases proibitórias ou concessórias às práticas envolvendo sistemas de IA (COMISSÃO EUROPEIA, 2022).

No documento da OCDE, o princípio aparece junto aos valores centrados no ser humano e na justi;ca (liberdade, dignidade e autonomia, privacidade, proteção de dados, não discriminação e igualdade, diversidade, justiça, justiça social e trabalho reconhecido internacionalmente (OCDE, 2019).

No Marco Civil, há uma seção (Seção II) dedicada à proteção dos dados pessoais. Nesse documento, o princípio aparece atrelado à preservação da intimidade, da vida privada, da honra e da imagem das partes envolvidas; isso, no entanto, não impede o acesso das autoridades administrativas aos dados cadastrais. De encontro ao que diz a atual Lei Geral de Proteção de Dados (LGPD), o artigo 11 menciona que operações envolvendo dados pessoais deverão respeitar o direito à privacidade e ao sigilo (BRASIL, 2014).

Na Portaria nº 271 de 04/12/2020 do CNJ o princípio não tem posição de ênfase e no PL 21/2020 e na Resolução nº 332 do CNJ aparece apenas como fundamento.

Na Resolução nº 395 do CNJ, o princípio aparece opostamente relacionado à transparência: "transparência: acesso à informação e aos dados produzidos pelo Poder Judiciário, respeitadas as hipóteses de restrição e de sigilo legal e a proteção de dados pessoais" (CNJ, 2021).

No PL 2.338, a proteção de dados aparece como fundamento (juntamente à privacidade e à autodeterminação informativa) e no tratamento de dados, que se guia pelas

medidas de privacidade por padrão e desde a concepção, a fim de minimizar o uso de dados pessoais (BRASIL, 2023).

5.3 DO PRINCÍPIO DA TRANSPARÊNCIA

A Resolução 2015/2103 (INL), sobre transparência, afirma que a transparência deve integrar a interoperabilidade entre sistemas abertos, a fim de evitar o bloqueio de sistemas exclusivos que limitem tal interoperabilidade. Na seção de princípios éticos, a transparência aparece juntamente à possibilidade de fundamentação de qualquer decisão tomada com recurso à IA que possa ter impacto substancial sobre a vida de indivíduos, também como forma de tornar a computação compreensível para os seres humanos, citando, ainda, que robôs avançados deveriam ser dotados de uma caixa preta com dados sobre as operações realizadas pela máquina, "incluindo os passos da lógica que conduziu à formulação das suas decisões". Por fim, na seção de licença para os criadores, aborda a garantia de se exigir a máxima transparência na programação de IA's, bem como a previsibilidade do comportamento robótico (PARLAMENTO EUROPEU, 2015).

Sobre a transparência, na seção Investigação e Inovação, o documento Comunicação “Inteligência artificial para a Europa”, explana que, para reforçar a confiança, as pessoas precisam compreender a tecnologia, de forma que "os sistemas de IA devem ser desenvolvidos de uma forma que permita ao ser humano ter um entendimento (das bases) das suas ações, a fim de aumentar a transparência e minimizar os riscos de distorção ou erro" (COMISSÃO EUROPEIA, 2018).

O documento “Plano Coordenado para a Inteligência Artificial” aborda a transparência juntamente à explicabilidade (COMISSÃO EUROPEIA, 2021).

Em *Building Trust in Human-Centric Artificial Intelligence,* o princípio da transparência aparece na rastreabilidade dos sistemas de IA, no registro, na explicabilidade (inclusive do grau de influência de orientação de um sistema de IA em determinado processo, da escolha de concepção, da justificativa da implementação do modelo, das limitações do sistema, etc) e na informação acerca da adequabilidade das capacidades e limitações do modelo (COMISSÃO EUROPEIA, 2019).

As Orientações Éticas para uma IA de Confiança citam explicitamente que o princípio da rastreabilidade "está estreitamente relacionado com o princípio da explicabilidade [tido como a capacidade de explicar tanto os processos técnicos de um sistema de IA como as decisões humanas com eles relacionadas] e abrange a transparência

dos elementos relevantes para um sistema de IA: os dados, o sistema e os modelos de negócio". A transparência estaria correlacionada também à comunicação, ou seja, para o documento, a necessidade de apresentar os sistemas de IA como máquinas e não como seres humanos, facilitando sua utilização (AI HLEG, 2019).

As Recomendações sobre a Ética da IA apresentam o princípio associado à explicabilidade dos sistemas inteligentes, como requisito essencial para a garantia e promoção de outros direitos humanos; o documento afirma que os dois aspectos devem ser adequados ao nível de contexto e impacto de cada situação, já que pode haver a necessidade de se equilibrar os dois princípios a outros (como a privacidade, a segurança e a proteção) (UNESCO, 2021).

A Declaração sobre Ética e Proteção de Dados em Inteligência Artificial traz o princípio da transparência relacionado ao da proteção de dados, como já mencionado na seção anterior; além disso, diferentemente do princípio da proteção de dados, a transparência é também relacionada à auditoria e à prestação de contas (responsabilidade) (ICDPPC, 2018).

A Carta Ética Europeia sobre o uso de IA nos sistemas judiciais cita o princípio da transparência junto ao da imparcialidade e equidade, a fim de tornar métodos de tratamento de dados acessíveis e compreensíveis (COMISSÃO EUROPEIA, 2018).

O documento "Digitalização para uma justiça melhor" menciona brevemente a transparência da justiça como meio de promoção da digitalização para melhorar o conhecimento sobre a justiça em geral, nomeadamente sobre a duração do processo (CEPEJ, 2021).

O Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica cita a transparência de regras de interoperabilidade, nos processos decisórios e dos sistemas algorítmicos (neste ponto, associa-a à explicabilidade, tratada como resultados "que sejam compreensíveis por públicos não técnicos e que lhes forneçam informações relevantes"). Na seção "Transparência, enviesamento e explicabilidade dos algoritmos", a carta traz o risco da tomada de decisões automatizadas no que se refere à estaticidade e à opacidade, enfatizando a necessidade de minimização da discriminação e parcialidade dos algoritmos através do desenvolvimento de um quadro ético comum sólido para o tratamento transparente de dados pessoais. O princípio aparece, ainda, relacionado à responsabilidade e à equidade, a fim de gerar um tratamento justo das orientações éticas em matéria de IA. Um dos pontos mais importantes do documento é o entendimento de que a divulgação do computador não resolve a questão

da transparência, porque não explica o processo de aprendizagem e que tal divulgação poderia, inclusive, conduzir ao uso indevido e à adulteração de algoritmos, desencorajando as empresas a desenvolver novos códigos já que a propriedade intelectual também estaria em risco. A transparência é intrinsecamente relacionada à inteligibilidade dos modelos inteligentes e pode revelar deficiências, mas não garante "a fiabilidade, a segurança e equidade", o que somente seria possível através de responsabilização (PARLAMENTO EUROPEU, 2019).

No Livro Branco, a transparência é trazida em oposição direta à opacidade e como requisito de promoção da utilização responsável da IA. Na carta, o princípio é relacionado à confiança, já que as informações sobre limitações e capacidades do modelo inteligente precisariam ser claras e precisas. O documento reproduz também algumas passagens de documentos elaborados pelo Grupo de Peritos de Alto Nível em IA (AI HLEG) e já citados neste trabalho anteriormente (COMISSÃO EUROPEIA, 2020).

Na Proposta de Regulamento sobre Inteligência Artificial, a transparência aparece relacionada à explicabilidade e à confiança. No que se refere aos modelos de risco, a carta traz o conceito de "transparência mínima", que permitiria que os indivíduos que interagissem com o conteúdo tomassem decisões informadas e tivessem a opção de continuar ou parar de usar a aplicação. O conceito de transparência utilizado no documento é o mesmo do "Draft AI R&D guidelines for international discussion", documento japonês de diretrizes para utilização de IA, que cita que "os desenvolvedores devem prestar atenção à verificabilidade das entradas/ saídas de sistemas de IA e a explicabilidade de seus julgamentos" (COMISSÃO EUROPEIA, 2022).

No documento da OCDE, o princípio é relacionado à explicabilidade como meio para alcançar a confiança. A tradução literal das recomendações traz:

> Os intervenientes na IA devem comprometer-se com a transparência e a divulgação responsável em relação aos sistemas de IA. Para tanto, eles devem fornecer informações significativas, apropriadas ao contexto e consistentes com o estado da arte: i. promover uma compreensão geral dos sistemas de IA, ii. conscientizar as partes interessadas sobre suas interações com os sistemas de IA, inclusive no local de trabalho, iii. para permitir que as pessoas afetadas por um sistema de IA compreendam o resultado e, iv. para permitir que aqueles afetados negativamente por um sistema de IA desafiem o seu resultado com base em informações de fácil compreensão sobre os fatores e a lógica que serviu de base para a previsão, recomendação ou decisão (OCDE, 2019).

O Marco Civil da Internet não traz menção expressa ao princípio; ao passo que a Portaria nº 271 de 04/12/2020 do CNJ só o menciona enquanto princípio geral.

A Resolução nº 332 do CNJ cita, em seu artigo 8º uma definição de transparência:

> I – divulgação responsável, considerando a sensibilidade própria dos dados judiciais; II – indicação dos objetivos e resultados pretendidos pelo uso do modelo de Inteligência Artificial; III – documentação dos riscos identificados e indicação dos instrumentos de segurança da informação e controle para seu enfrentamento; IV – possibilidade de identificação do motivo em caso de dano causado pela ferramenta de Inteligência Artificial V – apresentação dos mecanismos de auditoria e certificação de boas práticas; VI – fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial (CNJ, 2020).

A Resolução nº 395 do CNJ menciona a transparência como "acesso à informação e aos dados produzidos pelo Poder Judiciário, respeitadas as hipóteses de restrição e de sigilo legal e a proteção de dados pessoais" (CNJ, 2021).

O PL 21/2020, ao mencionar o princípio, juntamente com a o princípio da explicabilidade, afirma se tratarem da garantia sobre "o uso e funcionamento dos sistemas de inteligência artificial", além da "divulgação responsável do conhecimento de inteligência artificial, observados os segredos comercial e industrial, e de conscientização das partes interessadas sobre suas interações com os sistemas, inclusive no local de trabalho" (BRASIL, 2020).

O PL 2338 traz a transparência associada à explicabilidade, inteligibilidade e auditabilidade de processos. As medidas de transparência incluem o uso de "interfaces ser humano-máquina adequadas e suficientemente claras e informativas" e "transparência quanto às medidas de governança adotadas no desenvolvimento e emprego do sistema de inteligência artificial pela organização" (BRASIL, 2023).

5.4 DO PRINCÍPIO DA SUPERVISÃO HUMANA

Sobre a supervisão humana, a Resolução 2015/2103 (INL), apesar de não citar especificamente a expressão, traz, na introdução, a integração do controle e verificação humanos aos processos decisórios automatizados e algorítmicos. Também na seção "Princípios gerais relativos ao desenvolvimento da robótica e da inteligência artificial para utilização civil", afirma que

> [...] o desenvolvimento das tecnologias da robótica deve ser orientado para complementar as capacidades humanas, e não para as substituir; considera fundamental garantir que, no desenvolvimento da robótica e da inteligência artificial, os humanos tenham sempre o controlo sobre as máquinas inteligentes; considera que deve ser prestada particular atenção ao possível desenvolvimento de uma ligação emocional entre os seres humanos e os robôs, especialmente em grupos vulneráveis (crianças, idosos e pessoas com deficiência), e sublinha as questões suscitadas pelo grave impacto físico ou emocional que essa ligação emocional pode ter nos seres humanos (PARLAMENTO EUROPEU, 2015).

A Comunicação “Inteligência artificial para a Europa” não traz citações diretas ao princípio ou a princípios tratados como símiles (COMISSÃO EUROPEIA, 2018).

O “Plano Coordenado para a Inteligência Artificial” afirma que, no setor de saúde, por exemplo, a "IA pode revelar informações contidas nos dados para apoiar diagnósticos e tratamentos, mas deve ser sempre um médico humano a fazer as escolhas finais (supervisão humana)" (COMISSÃO EUROPEIA, 2021).

O documento "*Building Trust in Human-Centric Artificial Intelligence*" afirma que os sistemas de IA devem ajudar os indivíduos a fazerem escolhas de acordo com seus objetivos e não reduzir, limitar ou guiar a autonomia humana. Segundo a carta, "o controle humano contribui para garantir que um sistema de IA não prejudica a autonomia humana" e a supervisão pode ser realizada através de mecanismos de governação. A supervisão ou fiscalização humana ("human on the loop") é tratada por vezes como controle humano ("human in command") e por vezes como intervenção humana ("human in the loop"). Para a comissão, a intervenção deve ocorrer a cada ciclo de decisão do sistema e a supervisão refere-se à capacidade de intervenção e monitoração; já o conceito de controle refere-se à capacidade de supervisionar a atividade global do sistema de IA e de decidir como utilizá-lo (COMISSÃO EUROPEIA, 2019).

As Orientações Éticas para uma IA de Confiança afirmam que os sistemas de IA devem apoiar a autonomia e a tomada de decisão dos seres humanos, permitindo sua supervisão, utilizando-se dos mesmos conceitos da carta anterior (AI HLEG, 2019).

As Recomendações sobre a Ética da IA trazem a categorização da supervisão humana em individual e pública, argumentando que às vezes as pessoas podem decidir confiar nos modelos inteligentes por sua eficácia, mas a decisão de ceder o controle deve continuar sendo dos seres humanos, já que a prestação de conta e a responsabilidade finais

serão humanas. A fim de preservar a autonomia humana, no entanto, o documento sugere que os usuários possam solicitar intervenção humana (UNESCO, 2021).

A Declaração sobre Ética e Proteção de Dados em Inteligência Artificial afirma que os mecanismos de proteção de dados devem reforçar a supervisão dos sistemas de IA, estabelecendo mecanismos de transparência (ICDPPC, 2018).

A Carta Ética Europeia sobre o uso de IA nos sistemas judiciais afirma que técnicas de aprendizagem podem ou não ser supervisionadas por um humano; afirma que, sem substituir a intervenção humana, poderiam ser criados chatbots para facilitação do acesso à fontes de informação; cita o Regulamento (CE) n° 2016/679, que afirma que o titular de dados tem direito de obter intervenção humana por parte do responsável pelo tratamento de dados; e apresenta o princípio "sob controle do usuário", que visa impedir uma abordagem prescritiva e garantir que usuários sejam atores informados e controlem suas escolhas (COMISSÃO EUROPEIA, 2018).

O documento “Digitalização para uma justiça melhor” não traz citações expressas dos princípios (CEPEJ, 2021).

O Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica aborda que os titulares de dados têm direito à intervenção humana sempre que uma decisão se baseie num tratamento automatizado, leia-se, na íntegra:

> Sublinha que qualquer sistema de IA tem de ser desenvolvido no respeito dos princípios da transparência e da responsabilização relativamente aos algoritmos, de molde a permitir a compreensão das suas ações pelos seres humanos; observa que, para reforçar a confiança e permitir o progresso da IA, os utilizadores devem estar cientes da forma como os seus dados, bem como outros dados e dados inferidos a partir dos seus dados são utilizados quando comunicam ou interagem com um sistema de IA ou com seres humanos apoiados por um sistema de IA; considera que tal contribuirá para uma melhor compreensão e confiança por parte dos utilizadores; salienta que a inteligibilidade das decisões deve ser uma norma da UE, em conformidade com os artigos 13.º, 14.º e 15.º do Regulamento Geral sobre a Proteção de Dados (RGPD)[16]; recorda que o RGPD já prevê o direito de ser informado sobre a lógica subjacente ao tratamento de dados; sublinha que, de acordo com o artigo 22.º do RGPD, os titulares dos dados têm direito a uma intervenção humana sempre que uma decisão se baseie num tratamento automatizado que os afete significativamente (PARLAMENTO EUROPEU, 2019).

Em momento outro, o documento afirma que:

> Considerando que o desenvolvimento mais aprofundado e uma maior utilização de processos decisórios automatizados e algorítmicos têm um impacto inegável nas escolhas que uma pessoa a título individual (como, por exemplo, um homem de negócios ou um utilizador da Internet) e as autoridades administrativas, judiciárias ou outras autoridades públicas fazem para chegar a uma decisão final enquanto consumidores, empresas ou autoridades; que as garantias e a possibilidade de controlo e verificação humanos devem ser integradas nos processos decisórios automatizados e algorítmicos (PARLAMENTO EUROPEU, 2019).

De forma que os conceitos de controle (controlo) e verificação parecem se confundir à intervenção humana.

O Livro Branco traz a supervisão humana como um dos princípios elencados pelo Grupo de Alto Nível, afirmando que a norma garante que um sistema de IA não põe em causa a autonomia humana nem produz outros efeitos negativos, variando sua aplicabilidade em tipo e grau conforme o modelo inteligente:

> Por exemplo, a supervisão humana pode ter as seguintes manifestações, que não são exaustivas: o resultado do sistema de IA só se torna efetivo se tiver sido previamente revisto e validado por um ser humano (por exemplo, a decisão de rejeitar um pedido de prestações de segurança social só pode ser tomada por uma pessoa); o resultado do sistema de IA torna-se imediatamente efetivo, mas a intervenção humana é assegurada posteriormente (por exemplo, a rejeição de um pedido de cartão de crédito pode ser processada por um sistema de IA, mas a análise humana deve ser possível posteriormente); monitorização do sistema de IA durante o seu funcionamento e capacidade de intervenção em tempo real e de desativação (por exemplo, um botão ou um procedimento de paragem está 24 disponível num automóvel sem condutor quando um ser humano determina que o funcionamento do automóvel não é seguro); na fase de conceção, ao impor restrições operacionais ao sistema de IA (por exemplo, um veículo sem condutor deve deixar de funcionar em determinadas condições de baixa visibilidade quando os sensores podem tornar-se menos fiáveis ou manter uma determinada distância do veículo anterior em qualquer circunstância) (COMISSÃO EUROPEIA, 2020).

A Proposta de Regulamento sobre Inteligência Artificial (COMISSÃO EUROPEIA, 2022).

O documento da OCDE não cita expressamente nenhum dos princípios apresentados como símiles à supervisão humana (OCDE, 2019).

O Marco Civil e o PL 21/2020, além da Portaria nº 271 de 04/12/2020 e a Resolução nº 395, todos do CNJ, não trazem menções expressas aos princípios (BRASIL, 2014).

A Resolução nº 332 do CNJ menciona o controle do usuário sobre os dados, conceituando o usuário como a pessoa que "utiliza o sistema inteligente e que tem direito ao

seu controle". Além disso, a fim de assegurar a autonomia dos usuários internos, o documento enfatiza que os modelos inteligentes devem proporcionar incremento e não restrição à autonomia, possibilitando revisão da proposta de decisão e dos dados utilizados em sua elaboração, devendo permitir a supervisão do magistrado competente (CNJ, 2020).

O PL 2.338 traz a participação humana no ciclo da IA e a supervisão humana efetiva como princípios a serem seguidos no desenvolvimento, implementação e uso de sistemas inteligentes. A supervisão humana busca, para o documento, pode "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". Supervisão e controle podem ser dificultados pelo baixo grau de transparência, explicabilidade e auditabilidade da IA. Além disso, traz a possibilidade de solicitação de intervenção humana quando a decisão, previsão ou recomendação de modelo inteligente "produzir efeitos jurídicos relevantes ou que impactem de maneira significativa os interesses da pessoa, inclusive por meio da geração de perfis e da realização de inferências, esta poderá solicitar a intervenção ou revisão humana" (BRASIL, 2023).

5.5 DO PRINCÍPIO DA NÃO DISCRIMINAÇÃO

A Resolução 2015/2103 (INL) utiliza o termo genericamente e o atrela a outros princípios da carta, suscitando-o também na introdução, juntamente com o processo equitativo, a transparência e a inteligibilidade dos processos decisórios (PARLAMENTO EUROPEU, 2015).

A Comunicação “Inteligência artificial para a Europa” cita que, em função dos dados utilizados no treinamento da IA, seus resultados podem ser tendenciosos. Além disso, traz dados sobre a importância de se aumentar o número de profissionais da área de tecnologia da informação, sendo importante fomentar a diversidade, encorajando mais mulheres e pessoas de origens diversas, a fim de assegurar a inclusão (COMISSÃO EUROPEIA, 2018).

O “Plano Coordenado para a Inteligência Artificial” afirma que "é fundamental que exista uma interação eficaz entre as instalações de ensaio e experimentação e os espaços de dados para criar condições de concorrência equitativas e assegurar um acesso não discriminatório ao mercado", que são responsáveis pela fiabilidade e segurança das tecnologias de IA (COMISSÃO EUROPEIA, 2021).

O documento *Building Trust in Human-Centric Artificial Intelligence* afirma que os conjuntos de dados utilizados pela IA podem ser afetados pela inclusão inadvertida de modelos de má conduta, incompletos ou de má governação que geram vieses que poderiam conduzir à discriminação (in)direta - havendo, portanto, uma classificação tácita em discriminação direta e indireta. Para solucionar tais pontos, "os sistemas de IA devem ter em conta toda a gama de capacidades, competências e requisitos humanos", garantindo a acessibilidade a partir de uma concepção universal (COMISSÃO EUROPEIA, 2019).

As Orientações Éticas para uma IA de Confiança abordam o princípio da não discriminação juntamente à igualdade e à solidariedade em oposição à exclusão, afirmando que a não discriminação tolera o estabelecimento de distinções entre situações diferentes com base em justificações objetivas e que, num contexto de IA, a igualdade implica que as operações não gerem resultados injustamente tendenciosos. Para isso, a inclusão e a diversidade têm de estar presentes em todo o ciclo de vida do sistema de IA e o enviesamento identificável e discriminatório deve ser eliminado na fase de recolha de dados (AI HLEG, 2019).

As Recomendações sobre a Ética da IA afirmam que a não discriminação deve ser garantida a fim de garantir uma abordagem inclusiva, levando em consideração necessidades específicas de grupos etários, sistemas culturais, grupos linguísticos, pessoas com deficiências, etc, contribuindo para o ideal de justiça e igualdade (UNESCO, 2021).

A Declaração sobre Ética e Proteção de Dados em Inteligência Artificial afirma que muitos dos sistemas de IA acabam automatizando políticas discriminatórias contra populações em condições de vulnerabilidade (ICDPPC, 2018).

A Carta Ética Europeia sobre o uso de IA nos sistemas judiciais traz o princípio da não discriminação com o fim específico de prevenir o desenvolvimento ou intensificação de qualquer discriminação entre grupos ou indivíduos, seja através do agrupamento, da classificação de dados ou da utilização de quaisquer métodos que produzem ou reproduzem análises determinísticas e discriminatórias, devendo ser dispensada especial atenção ao tratamento de dados pessoais sensíveis. (COMISSÃO EUROPEIA, 2018).

A carta "Digitalização para uma justiça melhor" (CEPEJ, 2021) não traz menções diretas ao princípio.

O Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica afirma que a aprendizagem automática suscita desafios quanto à não discriminação, ao processo equitativo, à transparência e à inteligibilidade dos processos decisórios. Quando aborda o princípio, afirma que os cidadãos

não devem ser objeto de discriminação com base na sua classificação e que devem ter direito à outra oportunidade. À referência a tal classificação se encontra nos dados de treinamento da IA:

> Assinala que até dados de treino de elevada qualidade podem perpetuar a discriminação e a injustiça se não forem utilizados de forma cuidadosa e conscienciosa; observa que a utilização de dados de baixa qualidade, desatualizados, incompletos ou incorretos em diferentes fases do tratamento de dados pode conduzir a previsões e avaliações insuficientes e, por seu turno, a preconceitos, o que pode redundar em violações dos direitos fundamentais ou em conclusões puramente incorretas ou resultados falsos; considera, por conseguinte, que, na era dos megadados, é importante assegurar que os algoritmos sejam formados por amostras representativas de dados de elevada qualidade, de forma a alcançar a paridade estatística; realça que, mesmo que sejam utilizados dados exatos de elevada qualidade, uma análise preditiva baseada na IA só pode oferecer uma probabilidade estatística; recorda que, no âmbito do RGPD, o tratamento posterior de dados pessoais para fins estatísticos, inclusive o treino da IA, só pode gerar dados agregados que não podem voltar a ser aplicados a pessoas (PARLAMENTO EUROPEU, 2019).

O Livro Branco trabalha a não discriminação juntamente à diversidade e à equidade, categorizando tacitamente o princípio em discriminação proibida e não proibida:

> Requisitos para tomar medidas razoáveis destinadas a garantir que a utilização subsequente dos sistemas de IA não conduz a resultados que impliquem uma discriminação proibida. Estes requisitos podem implicar, em especial, a obrigação de utilizar conjuntos de dados suficientemente representativos, especialmente para assegurar que todas as dimensões relevantes de género, etnia e outros possíveis motivos de discriminação proibida sejam adequadamente refletidas nesses conjuntos de dados (COMISSÃO EUROPEIA, 2020).

A Proposta de Regulamento sobre Inteligência Artificial visa minimizar o risco de discriminação algorítmica, especialmente no que diz respeito "à conceção e à qualidade dos conjuntos de dados utilizados no desenvolvimento de sistemas de IA", bem como obrigações de testagem, gestão de riscos, documentação e supervisão humana ao longo do ciclo de vida dos sistemas de IA. Neste ponto, cita-se a disponibilidade de dados de elevada qualidade e práticas de governança e gestão de tais dados, para que sejam suficientemente e estatisticamente relevantes, representativos, livres de erros e completos, tendo em vista a finalidade prevista do sistema, com a devida observância de categorias especiais de dados pessoais (COMISSÃO EUROPEIA, 2022).

O documento da OCDE não traz menção expressa ao princípio (OCDE, 2019).

No Marco Civil, a não discriminação aparece no dever de tratamento isonômico a quaisquer pacotes de dados, sem distinção de conteúdo, origem, destino, serviço, terminal ou aplicação; o princípio é relacionado, no documento, à proporcionalidade, transparência e isonomia (BRASIL, 2014).

A Portaria nº 271 de 04/12/2020 e a Resolução nº 395, ambas do CNJ, não traz menção expressa ao princípio.

A Resolução nº 332 do CNJ traz a não discriminação juntamente à igualdade, à pluralidade, à solidariedade e ao auxílio no julgamento justo, criando condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos (CNJ, 2020).

O PL 21/2020 aborda o princípio da não discriminação junto à igualdade, à pluralidade e aos direitos trabalhistas, conceituando-o como a impossibilidade de uso dos sistemas para fins discriminatórios, ilícitos ou abusivos (BRASIL, 2020).

O PL 2.338 afirma que a não discriminação, juntamente à igualdade, à pluralidade e ao respeito aos direitos trabalhistas, são fundamentos do desenvolvimento, implementação e uso dos sistemas de IA no Brasil e conceitua como discriminação:

> [...] qualquer distinção, exclusão, restrição ou preferência, em qualquer área da vida pública ou privada, cujo propósito ou efeito seja anular ou restringir o reconhecimento, gozo ou exercício, em condições de igualdade, de um ou mais direitos ou liberdades previstos no ordenamento jurídico, em razão de características pessoais como origem geográfica, raça, cor ou etnia, gênero, orientação sexual, classe socioeconômica, idade, deficiência, religião ou opiniões políticas (BRASIL, 2023).

E como discriminação indireta:

> [...] discriminação que ocorre quando normativa, prática ou critério aparentemente neutro tem a capacidade de acarretar desvantagem para pessoas pertencentes a grupo específico, ou as coloquem em desvantagem, a menos que essa normativa, prática ou critério tenha algum objetivo ou justificativa razoável e legítima à luz do direito à igualdade e dos demais direitos fundamentais (BRASIL, 2023).

Diante do exposto, a seguir, traçam-se possíveis conclusões do arcabouço até aqui constituído.

5.6 PONDERAÇÕES E TENDÊNCIAS ACERCA DO ARCABOUÇO CONSTITUÍDO

Inicialmente, cabe mencionar que, por vezes, os princípios sequer são mencionados nas cartas. Quando o são, raramente são conceituados.

Da análise de recorte do arcabouço teórico, é possível inferir uma preocupação severa com uma "IA responsável", com o homem no centro do processo (também chamada "abordagem centrada no ser humano" ou "centralidade no ser humano"). A autonomia humana (que também é tratada como controle humano - como na Carta Ética Europeia sobre o uso de IA nos sistemas judiciais, ou como intervenção humana - como pela OCDE) aparece, portanto, como valor caro às boas práticas na utilização da IA, o que também é demonstrado na máxima de "capacitar as pessoas e não as substituir".

Insta destacar que a confiança nos modelos inteligentes parece contar com valores como a cibersegurança e a proteção de dados, devendo assegurar, ao mesmo tempo, conforme documentos como “Aumentar a confiança numa inteligência artificial centrada no ser humano”, a qualidade dos dados.

Ponto de especial ênfase são normativas especializadas em IA que não apresentam aprofundamento técnico específico. Ou seja, quando o Livro Branco Sobre a Inteligência Artificial apresenta questões como a apresentação de dados de treino e a utilização de "requisitos específicos para determinadas aplicações de IA"; quando a Proposta de Regulamento sobre Inteligência Artificial abarca questões como a tradução de "princípios éticos de alto nível em critérios verificáveis que ajudam a moldar o design, desenvolvimento, implantação e uso da IA ética"; ou quando o PL 21/2020 afirma que deve haver uma "busca pela neutralidade", parece haver pouca clareza de quais técnicas, de fato, os especialistas deveriam utilizar.

Ainda sobre tal imbróglio, o Direito parece trazer propostas como a "adoção de tecnologias, padrões e formatos abertos e livres", a "disponibilização dos dados utilizados para treinamento no modelo" (Portaria nº 271 de 04/12/2020), pouco integradas aos modelos de negócio já existentes, desconsiderando as informações sigilosas de empresas, por exemplo.

Sobre o princípio da responsabilidade, mais especificamente, alguns documentos propõem uma aplicação objetiva pautada na gestão de riscos e de acordo com o nível efetivo de instruções dadas ao robô e à sua autonomia (Resolução 2015/2103, INL). Outras fontes apresentam a maior importância da auditabilidade e da possibilidade de reparação (*Building*

Trust in Human-Centric Artificial Intelligence); outras, do controle ou supervisão humanos na relação com tal princípio.

Também há documentos que tratam a responsabilidade como sinônima de *accountability* e de prestação de contas, havendo divergência acerca da responsabilização. As Recomendações Sobre a Ética da IA traz uma aproximação entre responsabilidade e prestação de contas, sem, no entanto, tratar os princípios como sinônimos. Já a Portaria nº 271 de 04/12/2020 afirma que a responsabilidade sobre proteção dos dados é ônus do criador e/ou mantenedor do modelo inteligente; no PL 21/2020, os agentes de desenvolvimento e operação são responsáveis; no PL 2.338, a atribuição da responsabilidade aparece no desenvolvimento, implementação e uso dos sistemas de IA.

Sobre a proteção de dados, o princípio é tratado por vezes como sinônimo de privacidade (como é o caso do documento Orientações Éticas para uma IA de Confiança), aparecendo em oposição à transparência (Resolução nº 395 do CNJ). Outros documentos, como a Declaração sobre Ética e Proteção de Dados em Inteligência Artificial, afirmam que a ausência de transparência de coleta de dados ou de tomada de decisões das máquinas geram tensões no sistema tradicional de proteção de dados.

Sobre a transparência, é possível identificar coerência entre os documentos na relação do princípio com o entendimento humano, a explicabilidade, a inteligibilidade e não somente a divulgação. Se, de um lado, documentos como a Carta Ética Europeia sobre o uso de IA nos sistemas judiciais cita o princípio da transparência juntamente à imparcialidade e à equidade; o Relatório sobre uma Política Industrial Europeia Completa no Domínio da Inteligência Artificial e da Robótica, por exemplo, afirma que a equidade do sistema só é possível mediante responsabilização.

Sobre a supervisão humana, tal expressão é relacionada ao controle, à autonomia, à fiscalização, à validação, à análise, ao monitoramento, à verificação e à intervenção humanos, havendo pouca ou nenhuma distinção entre os termos *human on the loop, human in the loop* e *human in command* entre as cartas.

Sobre a não discriminação, há menções, de um lado, sobre a ausência de diversidade e a necessidade de inclusão, como no documento Comunicação "Inteligência artificial para a Europa" e nas Orientações Éticas para uma IA de Confiança, que afirmam que a não discriminação tolera o estabelecimento de distinções entre situações diferentes com base em

justificações objetivas e que, num contexto de IA, a igualdade implica que as operações não gerem resultados injustamente tendenciosos. Por outro lado, no entanto, documentos como o Marco Civil afirmam que o dever na utilização ética da IA parte de um tratamento isonômico, sem distinção.

CONCLUSÕES

Diante do exposto, é possível inferir que existem diversas dificuldades de compatibilização principiológica no arcabouço constituído.

Inicialmente porque os princípios, por vezes, sequer são mencionados em alguns documentos, o que aponta para um dissenso no processo de eleição de priorização; além disso, o elevado número de princípios encontrados na análise de *corpus* apresenta um cenário parecido, em que há diversas normas, por vezes compatíveis, por vezes conflitantes, utilizadas a um só tempo.

Em segundo lugar, quando citados, os princípios não são conceituados, talvez por conta da importação principiológica de outro quadro jurídico, que tampouco é mencionado; os documentos se limitam a mencionar a interligação daquele princípio com outro do quadro. Nenhum é definido pela carta e as correlações aparentemente são realizadas sem demonstração de critério, tanto que as concatenações são diversas e distintas entre os documentos.

Há, ainda, em terceiro lugar, a dificuldade da linguagem. Porque os quadros jurídicos são importados, adiciona-se o elemento da tradução e um mesmo princípio pode aparecer com diferentes denominações. Proteção de dados por vezes é tratada como privacidade de dados; solidez, que é uma palavra que aparece muito no contexto principiológico da utilização da IA (como em solidez técnica), por vezes é traduzido como robustez; equidade pode aparecer como igualdade; prestação de contas, responsabilidade, responsabilização e *accountability* são tratadas como palavras símiles; o princípio da supervisão humana ora é traduzido como supervisão, ora como princípio do controle humano, ora como agência humana, ora autonomia humana; neutralidade e imparcialidade podem aparecer como sinônimos; não causar dano e não maleficência também.

Em quarto lugar, há o problema interpretativo, hermenêutico, que parece ser o primeiro que salta aos olhos durante as análises: uma das principais dificuldades encontradas na análise dos diferentes princípios levantados é a das diferentes interpretações para um mesmo conceito na literatura específica. Um mesmo princípio pode ser entendido a partir de duas perspectivas, como é o caso do princípio da não discriminação, que, por vezes, é tratada como igualdade no tratamento e, por vezes, como especificidade diante de cada tipo de dado. O princípio da transparência é por vezes interpretado como inteligibilidade e por vezes como disponibilização da informação. É possível perceber, por exemplo, que a privacidade e a proteção de dados são tratados como princípios cujas

definições são equiparadas, por exemplo. Na seara da *accountability*, sequer há consenso sobre a definição. Isso faz com que não exista consistência entre conceituações dos princípios, que às vezes são utilizados como sinônimos de coisas distintas ou até mesmo opostas.

Por fim, há o problema da integração do direito à tecnologia. Porque existem lacunas normativas na seara dos dados e da inteligência artificial, especialmente diante da contínua inovação tecnológica, e porque não há consenso acerca do quadro normativo a ser utilizado, principalmente por conta da especificidade situacional, a adequação técnica do Direito torna-se inócua. Há especificações que contradizem valores já existentes no Direito, como o de sigilo de negócio, e, por vezes, sequer há especificação técnica sobre o que deve ser de fato performado pelos especialistas da ponta tecnológica, havendo uma redução jurídica tecnicista e abstrata a frases generalistas como "deve-se cumprir o direito à transparência".

As dificuldades de aplicação dos princípios não interferem apenas na efetividade das normativas, mas em imbróglios como o da motivação decisional. Se, no início deste estudo, a preocupação era acerca da inescrutabilidade de decisões tomadas por agentes inteligentes, ao fim, a preocupação se replica na ausência de rigor conceitual no momento da eleição de um ou outro princípio, o que transforma a motivação humana numa verdadeira "bolsa de gatos", impossibilitando o amplo debate sobre decisões judiciais e a efetivação da justiça e tornando as decisões tomadas por agentes humanos tão inescrutáveis quanto aquelas tomadas por agentes inteligentes.

REFERÊNCIAS

AI HLEG. Orientações Éticas Para Uma Ia De Confiança. **Grupo Independente De Peritos De Alto Nível Sobre A Inteligência Artificial**, 2019. Disponível em: https://escola.mpu.mp.br/servicos-academicos/atividades-academicas/inovaescola/atividades-de-extensao/3-ciclo-de-debates/inteligencia-artificial-e-internet-das-coisas-oportunidades-e-desafios/ethicsguidelinesfortrustworthyai-ptpdf.pdf.

ALEXY, Robert. **Teoria da Argumentação Jurídica**. 2. ed. São Paulo: Landy, 2005.

ALVES, Rachel Cristina Vesú. **Metadados como elementos do processo de catalogação**. Tese (Doutorado em Ciência da Informação) **- Faculdade de Filosofia e Ciências**, Universidade Estadual Paulista, Marília, SP, 2010.

AMARAL, Rogério do. **Exposição privada nas redes sociais: uma análise sobre o Facebook na sociedade contemporânea**. Tese (Doutorado). UNESP, FCT, 2016.

ARAÚJO, Franciele Cassimiro de; ROSSI, Jackeline Magrin. **A evolução dos ataques cibernéticos**. Americana, 2020.

BARROS, Frederico Kern Ferreira. **Fake News, legislação simbólica e a proteção dos direitos fundamentais e da personalidade digital**. In.: OMMATI, José Emílio Medauar (org.). Escritos de direitos fundamentais. Belo Horizonte: Conhecimento Editora, 2021.

BENBASAT, I.; GOLDSTEIN, D. K.; MEAD, M. The case research strategy in studies of information systems. **MIS Quarterly**, v. 11, n. 3, p. 369-386, set. 1987.

BIONI, Bruno. Compreendendo o conceito de anonimização e dado anonimizado. **Cadernos Jurídicos**. São Paulo, v. 21, p. 191-201, 2020.

BIONI, Bruno. **Proteção de Dados Pessoais**: a função e os limites do consentimento. Rio de Janeiro: Forense, 2019.

BOURDIEU, Pierre. **A distinção**. São Paulo: Edusp, 2007.

BRASIL. **Decreto nº 8.771, de 11 de maio de 2016**. Regulamenta a Lei nº 12.965, de 23 de abril de 2014 para tratar das hipóteses admitidas de discriminação de pacotes de dados na internet e de degradação de tráfego, indicar procedimentos para guarda e proteção de dados por provedores de conexão e de aplicações, apontar medidas de transparência na requisição de dados cadastrais pela administração pública e estabelecer parâmetros para fiscalização e apuração de infrações. Diário Oficial União, 2016.

BRASIL. **Decreto nº 8.777, de 11 de maio de 2016**. Institui a Política de Dados Abertos do Poder Executivo federal. Diário Oficial da União, 2016.

BRASIL. **Emenda Constitucional nº 115, de 10 de fevereiro de 2022.** Altera a Constituição Federal para incluir a proteção de dados pessoais entre os direitos e garantias fundamentais e para fixar a competência privativa da União para legislar sobre proteção e tratamento de dados pessoais. Diário Oficial da União, 2022.

BRASIL. **Lei nº 12.737, de 30 de novembro de 2012**. Dispõe sobre a tipificação criminal de delitos informáticos; altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 - Código Penal; e dá outras providências. Diário Oficial da União, 2012.

BRASIL. **Lei nº 12.965, de 23 de abril de 2014**. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Diário Oficial da União, 2014.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União, 2018.

BRASIL. **Portaria nº 271, de 04 de dezembro de 2020**. Regulamenta o uso de Inteligência Artificial no âmbito do Poder Judiciário. Conselho Nacional de Justiça, 2020.

BRASIL. **Projeto de Lei n° 21, de 07 de julho de 2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Câmara dos Deputados, 2020. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/151547.

BRASIL. **Projeto de Lei n° 2338, de 21 de fevereiro de 2023**. Dispõe sobre o uso da Inteligência Artificial. Câmara dos Deputados, 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233.

BRASIL. **Resolução nº 332, de 21 de agosto de 2020**. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. **Conselho Nacional de Justiça**, 2020.

BRASIL. **Resolução nº 395, de 7 de junho de 2021**. Institui a Política de Gestão da Inovação no âmbito do Poder Judiciário. Conselho Nacional de Justiça, 2021.

BUTCHER, Isabela. Novas leis europeias de serviços e mercados digitais terão impacto global. **Mobile Time**, 2022. Disponível em https://www.mobiletime.com.br/noticias/07/07/2022/o-impacto-global-das-novas-leis-de-mercado-e-servico-na-ue/. Acesso em 23 de set. de 2022.

CAMPOS, Ana Maria. Accountability: quando poderemos traduzi-la para o português? Revista de Administração Pública. Rio de Janeiro, **FGV**, 24(2):30-50, fev./abr. 1990.

CASTELLS, Manuel. **A sociedade em rede**. Revista e ampliada. Tradução de Roneide Majer. São Paulo: Paz e Terra, 2016.

CEPEJ. **Digitalisation for a better justice**. CEPEJ Action plan, 2021. Disponível em: https://rm.coe.int/cepej-2021-12-en-cepej-action-plan-2022-2025-digitalisation-justice/1680a4cf2c.

CHAWLA, Ronit. Deepfakes: How a pervert shook the world. **International Journal of Advance Research and Development**, v. 4, n. 6, p. 4-8, 2019.

COMISSÃO EUROPEIA. **Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu Ambiente**. Bruxelas: Comissão Europeia**,** 2018. Disponível em: https://rm.coe.int/carta-etica-traduzida-para-portugues-revista/168093b7e0.

COMISSÃO EUROPEIA. **Communication**: Building Trust in Human Centric Artificial Intelligence. Bruxelas: Comissão Europeia, 2019. Disponível em: https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence.

COMISSÃO EUROPEIA. **Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões Inteligência Artificial para a Europa**. Bruxelas: Comissão Europeia, 2018.

COMISSÃO EUROPEIA. **Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité Das Regiões**. Aumentar a confiança numa inteligência artificial centrada no ser humano. Bruxelas: Comissão Europeia, 2019. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52019DC0168.

COMISSÃO EUROPEIA. Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões: Pacto Ecológico Europeu. **Pacto Ecológico Europeu**, v. 11, p. 2019, 2019.

COMISSÃO EUROPEIA. **Coordinated Plan on Artificial Intelligence**. Bruxelas: Comissão Europeia, 2021.

COMISSÃO EUROPEIA. **Livro Branco sobre a inteligência artificial**: uma abordagem europeia virada para a excelência e a confiança. Bruxelas: Comissão Europeia, 2020.

COMISSÃO EUROPEIA. **Proposta de Regulamento sobre Inteligência Artificial**. Bruxelas: Comissão Europeia, 2022. Disponível em: https://repositorio.enap.gov.br/bitstream/1/7419/1/2022.12.08%20-%20Regula%C3%A7%C3%A3o%20da%20Intelig%C3%AAncia%20Artificial.pdf.

DAL BELLO, Cíntia. Visibilidade, vigilância, identidade e indexação: a questão da privacidade nas redes sociais digitais. **Logos**, v. 18, n. 1, 2011.

DONEDA, Danilo. A proteção dos dados pessoais como um direito fundamental. **Espaço Jurídico**, Joaçaba, v. 12, n. 2, p. 91-108, jul/dez 2011. Disponível em: http://editora.unoesc.edu.br/index.php/espacojuridico/article/download/1315/658. Acesso em: 01 set. 2017.

DOS SANTOS GOMES, Dennis. Inteligência Artificial: Conceitos e Aplicações. **Revista Olhar Científico**, v. 01, n. 02, ago/dez 2010.

EBIA. **Estratégia brasileira de inteligência artificial**. Ministério da Ciência, Tecnologia e Inovações Secretaria de Empreendedorismo e Inovação, 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf.

ENGELMANN, Wilson; MARQUES, Clarice Gonçalves Pires. Inteligência Artificial e as Configurações Contemporâneas do Direito: da Inovação Tecnocientífica à Inovação Justecnológica. **Revista de Direito Brasileira**, v. 28, n. 11, p. 405-421, 2021.

EUROPA. **Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões**. Inteligência artificial para a Europa {SWD(2018) 137 final}. Comissão Europeia: Bruxelas, 2018.

FORTINET. Conheça os 11 tipos mais comuns de ataques cibernéticos a empresas e descubra como se proteger. **Sigma Telecom**, 2022. Disponível em https://www.sigmatelecom.com.br/11-tipos-comuns-de-ataques-ciberneticos/. Acesso em 11 de set. de 2022.

FOUCAULT, Michel. **Arqueologia do saber**. 6 ed. Rio de Janeiro: Forense Universitária, 2002.

FRAZÃO, Ana. **Algoritmos e inteligência artificial**. Jota, 2018.

FRAZÃO, Ana. Fundamentos da proteção dos dados pessoais: Noções introdutórias para a compreensão da importância da Lei Geral de Proteção de Dados. **Lei Geral de Proteção de Dados Pessoais e suas repercussões no Direito Brasileiro**, v. 1, p. 23-52, 2019.

G1. Denúncias de crimes cometidos pela internet mais que dobram em 2020. **G1**, 2021. Disponível em: https://g1.globo.com/economia/tecnologia/noticia/2021/02/09/numero-de-denuncias-de-crimes-cometidos-pela-internet-mais-que-dobra-em-2020.ghtml. Acesso em: 11 de set. de 2022.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. O'Reilly Media, Inc., 2022.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. São Paulo: Atlas, 1995.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2008.

GONZÁLEZ. Conheça o cenário das leis de proteção de dados ao redor do mundo. **Blog Id Wall**, 2020. Disponível em https://blog.idwall.co/protecao-de-dados-cenario-mundial-das-leis/. Acesso em 23 de set. de 2022.

HAN, Jiawei; PEI, Jian; TONG, Hanghang. **Data mining**: concepts and techniques. Morgan kaufmann, 2022.

HART, C. **Doing a Literature Review**: Releasing the Social Science Research Imagination. London: SAGE Publications, 1998.

HERN, Alex. **'Anonymous' browsing data can be easily exposed, researchers reveal**. The Guardian, 2017. Disponível em http://www.theguardian.com/technology/2017/aug/01/data-browsinghabits-broker. Acesso em 08 de set. de 2022.

HOUAISS, Antônio. **Dicionário Houaiss da língua portuguesa**. Disponível em https://houaiss.uol.com.br/corporativo/apps/uol_www/v6-0/html/index.php#2. Acesso em 08 de set. de 2022.

HOULIHAN, David. ROSS**. Intelligence & Artificial Intelligence in Legal Research**. Blue Hill Research, 2017.

ICDPPC, International Conference of Data Protection and Privacy Commissioners. **Declaration on Ethics and Data Protection in Artificial Intelligence**, 2018. Disponível em: https://icdppc.org/wp-content/uploads/2018/10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf.

JÚNIOR, Antonio Rulli; NETO, Antonio Rulli. Direito ao Esquecimento e o Superinformacionismo: apontamentos no direito brasileiro dentro do contexto de sociedade da informação. **Revista ESMAT**, v. 5, n. 6, p. 11-30, 2013.

KASPERSKY. Ciberataques crescem 23% no Brasil em 2021. **Kaspersky**, 2021. Disponível em: https://www.kaspersky.com.br/blog/panorama-ciberameacas-brasil-2021-pesquisa/18020/. Acesso em: 11 de set. de 2022.

KLEINBERG, J.; EASLEY, D. **Networks, Crowds, and Markets**: Reasoning about a Highly Connected World. Cambridge: University Press, 2010.

LAMBRECHT, Anja; TUCKER, Catherine. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. **Management science**, v. 65, n. 7, p. 2966-2981, 2019.

LAVAL, Christian; DARDOT, Pierre. **A nova razão do mundo**: ensaio sobre a sociedade neoliberal. São Paulo: Boitempo, 2016.

LÉVY, P. **Cibercultura**. São Paulo: Editora 34, 1999.

LIN, Jie et al. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. **IEEE Internet of things journal**, v. 4, n. 5, p. 1125-1142, 2017.

LIRA, L. de A. **Lei Carolina Dieckmann**: (in) eficácia da proteção dos direitos fundamentais à intimidade e à vida privada em face da pena cominada aos delitos informáticos. Brasília, DF: Conteúdo Jurídico, 2014.

LOURENÇO, Gabriel D. Cibercrimes no Brasil crescem em 23% em 2021, aponta pesquisa. **Olhar Digital**, 2022. Disponível em: https://olhardigital.com.br/2021/08/31/seguranca/cibercrime-brasil-2021/. Acesso em: 11 de set. de 2022.

MAFFESOLI, M. Perspectivas tribais ou a mudança do paradigma social. **Revista FAMECOS**, Porto Alegre, n. 23, abr. 2004.

MANZANO, José Augusto NG; DE OLIVEIRA, Jayr Figueiredo. **Algoritmos lógica para desenvolvimento de programação de computadores**. Saraiva Educação SA, 2000.

MARAS, Marie-Helen; ALEXANDROU, Alex. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. **The International Journal of Evidence & Proof**, v. 23, n. 3, p. 255-262, 2019.
MATSUBARA, Edson Takashi. **O algoritmo de aprendizado semi-supervisionado co-training e sua aplicação na rotulação de documentos**. 2004. Tese de Doutorado. Universidade de São Paulo.

MCCARTHY, J., MINSKY, Marvin L., ROCHESTER Nathaniel, SHANNON, Claude E. **Uma proposta para o Projeto Dartmouth Summer Pesquisa sobre Inteligência Artificial**. OPENCADD, 1955.

MCCARTHY, John. Generality in artificial intelligence. **Communications of the ACM**, v. 30, n. 12, p. 1030-1035, 1987.

MCCARTHY, John. **What is artificial intelligence**. 2007.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, p. 115-133, 1943.

MONTREAL UNIVERSITY. **The Montreal Declaration for a Responsible Development of Artificial Intelligence**: a participatory process. Universidade de Montreal, 2017. Disponível em:< https://www. montrealdeclaration-responsibleai.com/>. Acesso em 30 de janeiro de 2019.

MORENO, Guillermo Palao. A União Europeia dá seus primeiros passos na regulamentação da relação entre inteligência artificial e propriedade intelectual. **Revista Rede de Direito Digital, Intelectual & Sociedade**, v. 1, n. 1, p. 45-68, 2021.

MOSES, Lyria Bennet**. How to Think about Law, Regulation and Technology**: Problems with 'Technology' as a Regulatory Target. SSRN-Elsevier, 2014.

MURDOCK, Graham. Media materialties: For a moral economy of machines. **Journal of Communication**, v. 68, n. 2, p. 359-368, 2018.

NEGRI, Sergio Marcos Carvalho de Ávila; PELUSO LOPES, Giovana F. Da personalidade eletrônica à classificação de riscos na Inteligência Artificial (IA). **Teoria Jurídica Contemporânea**, v. 6, n. 1, 2021.

NUNES, Ana Carolina de Assis. **Entre redes neurais naturais e artificiais**: estudo antropológico sobre humanidade e inteligência artificial em algumas revistas brasileiras. Dissertação de Mestrado. Universidade Federal de Goiás, Faculdade de Ciências Sociais, Programa de Pós-Graduação em Antropologia Social.

O DILEMA DAS REDES. **Netflix**, 2020. Documentário disponível em: https://www.netflix.com. Acesso em 11 de set. de 2022.

O GUIA FINANCEIRO. **Big Data**: Como a Target descobriu uma gravidez antes da família?. O Guia Financeiro, 2019. Disponível em https://www.oguiafinanceiro.com.br/textos/big-data-como-a-target-descobriu-uma-gravidez-antes-da-propria-familia/. Acesso em 11 de set. de 2022.

OCDE, Organisation for Economic Co-operation and Development. **Recommendation of the Council on Artificial Intelligence**, 2019. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

O'NEIL, Cathy. **Algoritmos de destruição em massa**. Editora Rua do Sabão, 2021.

PARLAMENTO EUROPEU. **General Data Protection Regulation**. Bruxelas, 2016.

PARLAMENTO EUROPEU. **Relatório que contém recomendações à Comissão sobre Disposições de Direito Civil sobre Robótica**. Parlamento Europeu, 2017.

PARLAMENTO EUROPEU. **Resolução do Parlamento Europeu de 16 de fevereiro de 2017, com recomendações à Comissão de Direito Civil sobre Robótica** (2015/2103(INL)). PRESS RELEASES, 2017.

PARLAMENTO EUROPEU. **Robots and artificial intelligence**: MEPs call for EU-wide liability rules. Parlamento Europeu. **Press Releases**, 2017. Disponível em: https://www.europarl.europa.eu/news/en/press-room/20170210IPR61808/robots-and-artificial-intelligence-meps-call-for-eu-wide-liability-rules. Acesso em: 18 set. 2021.

PIERRE LEVY. **Cibercultura**. Editora 34, 2010.

PORTAL DA INDÚSTRIA. INDÚSTRIA DE A - Z. Indústria 4.0: Entenda seus conceitos e fundamentos. **Portal Da Indústria**, 2020. Disponível em: https://www.portaldaindustria.com.br/industria-de-a-z/industria-4-0/. Acesso em 3 de out. de 2022.

QIN, Jian; ZENG, Marcia Lei. **Metadata**. American Library Association, 2020.

RIBEIRO, Marco Tulio; SINGH, Sameer; GUESTRIN, Carlos. "Why should i trust you?" Explaining the predictions of any classifier. In: **Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining**. 2016. p. 1135-1144.

ROCHER, Luc; MUTHU, Meenatchi Sundaram; DE MONTJOYE, Yves-Alexandre. The observatory of anonymity: An interactive tool to understand re-identification risks in 89 countries. In: **Companion Proceedings of the Web Conference 2021**, 2021.

RODOTÀ, Stefano et al. A vida na sociedade da vigilância: a privacidade hoje. In: **A vida na sociedade da vigilância: a privacidade hoje**, 2008.

ROETS, Arne et al. ‘Fake news’: Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. **Intelligence**, v. 65, p. 107-110, 2017.

RUSSELL, Stuart J. **Artificial intelligence a modern approach**. Pearson Education, Inc., 2021.

SAIKALI, Lucas Bossoni. O Dilema de Collingridge e as novas tecnologias: quando regular? **Jornal de Direito Administrativo (JDA)**, ISSN 2675-2921, a. 1, v. 1, n. 4, 2020.

SALOMÃO, Luiz Felipe**. Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário brasileiro**. Rio de Janeiro: Centro de Inovação, Administração e Pesquisa do Judiciário da Fundação Getúlio Vargas, 2020.

SAYURI, Juliana. Retratos de uma juventude. **Estadão**, 2014. Disponível em: http://www.estadao.com.br/noticias/geral,retrato-de-umajuventude,1167792. Acesso em: 11 de set. de 2022.

SCAPINI, Luísa Almeida Ribeiro. **Responsabilidade civil extracontratual por danos causados por sistemas de inteligência artificial:** soluções disruptivas no Direito. 2020. Dissertação de Mestrado.

SCHNEIDER, Susan (Ed.). **Science fiction and philosophy**: from time travel to superintelligence. John Wiley & Sons, 2016.

SCHWAB, Klaus. **A quarta revolução industrial**. Edipro, 2019.

SEAVER, N. Knowing Algorithms. In: VERTESI, J.; RIBES, D. (Eds.). DigitalSTS: A Field Guide for Science & Technology Studies. Princeton: **Princeton University Press**, 2019.

SILVA, Ivan Nunes da; SPATTI, Danilo Hernane. **Redes Neurais Artificiais**: Teoria e Aplicações em Inteligência Computacional. 2ª ed. São Paulo: Blucher, 2017.

SILVA, Rodrigo Guedes. **Introdução da Inteligência Artificial Aplicada no Processo de Tomada de Decisões no Poder Judiciário Brasileiro**, 2022.

SILVEIRA, Sergio Amadeu. Discursos sobre regulação e governança algorítmica. **Estudos de sociologia**, v. 25, n. 48, 2020.

SILVEIRA, Sérgio Amadeu. Responsabilidade algorítmica, personalidade eletrônica e democracia. **Revista Eletrônica Internacional de Economia Política da Informação, da Comunicação e da Cultura**, v. 22, n. 2, p. 83-96, 2020.

SIQUEIRA, Alessandra Cristina de Mendonça. O Colonialismo Digital Como Nova Forma de Imperialismo na Sociedade em Rede. **Revista do Mestrado em Direito da UFS**, V.8, N.01, p. 29 – 50, Jan-Jun/2019, ISSN 2237-2040, 2019.

SRNICEK, N. **Platform Capitalism**. New York: John Wiley & Sons, 2016.

TIMPONE, GUIDI; GUIDI, Michel. **Explorando a Mudança de Cenário da IA**. Da IA Analítica à IA Generativa, p. 2023-05, 2023.

TIWARI, Tanya; TIWARI, Tanuj; TIWARI, S. How Artificial Intelligence, Machine Learning and Deep Learning are Radically Different? **International Journal of Advanced Research in Computer Science and Software Engineering**, v. 8, n. 2, p. 1-9, mar. 2018. ISSN 2277128X. Disponível em: http://ijarcsse.com/index.php/ijarcsse/article/view/569. Acesso em: 07 set. 2019.

TURING, Alan M. **Computing machinery and intelligence**. Springer Netherlands, 2009.
UNESCO. **Recomendação sobre a Ética da Inteligência Artificial,** 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por.

UNIÃO EUROPEIA. Proposta de regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial. **Regulamento Inteligência Artificial**. Altera determinados atos legislativos da União. Bruxelas: Comissão Europeia. Disponível em: https://eur-lex. europa. eu/legal-content/PT/TXT/HTML, 2021.

VAN DIJCK, José; POELL, Thomas; DE WAAL, Martijn. **The platform society**: Public values in a connective world. Oxford university press, 2018.

VENTURA, Katia Santiago; DE ALBUQUERQUE SIEBRA, Sandra; LIMA, Marcos Galindo. **A visão sistêmica na efetivação da Lei de Acesso à Informação**: o caso da Universidade Federal de Pernambuco. UFPE, 2013.

VIEIRA, Leonardo Marques. A problemática da inteligência artificial e dos vieses algorítmicos: caso COMPAS. In: **Brazilian Technology Symposium**. 2019.

VIEIRA, Tatiana Malta. **O direito à privacidade na sociedade da informação**: efetividade desse direito fundamental diante dos avanços da tecnologia da informação. Porto Alegre: Sergio Fabris, 2007.

ZENO-ZENCOVICHI. Legal epistemology in the times of Big Data. In: Knowledge of the law in the Big Data Age. Ginevra Peruginelli e Sebastiano Faro (ed). Netherlands: **IOS Press Bv**, 2019, p. 3.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância**. Editora Intrinseca, 2021.