UNIVERSIDADE FEDERAL DE ALAGOAS - UFAL
FACULDADE DE DIREITO DE ALAGOAS - FDA
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO PÚBLICO - PPGD

IZAAC DUARTE DE ALENCAR

A PROTEÇÃO DE DADOS PESSOAIS E INTELIGÊNCIA ARTIFICIAL: TRATAMENTO AUTOMATIZADO E O DIREITO À EXPLICAÇÃO

MACEIÓ-AL
2023

IZAAC DUARTE DE ALENCAR

A PROTEÇÃO DE DADOS PESSOAIS E INTELIGÊNCIA ARTIFICIAL: TRATAMENTO AUTOMATIZADO E O DIREITO À EXPLICAÇÃO

Dissertação de Mestrado apresentada no Programa de Pós-Graduação em Direito Público da Universidade Federal de Alagoas, como requisito parcial para obtenção do grau de Mestre em Direito.

Orientadora: Profª. Drª. Juliana de Oliveira Jota Dantas

Maceió-AL
2023

Catalogação na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

A368p Alencar, Izaac Duarte de.
A proteção de dados pessoais e inteligência artificial : tratamento automatizado e o direito à explicação / Izaac Duarte de Alencar. – 2023.
161 f.

Orientadora: Juliana de Oliveira Jota Dantas.
Dissertação (mestrado em Direito) – Universidade Federal de Alagoas. Faculdade de Direito de Alagoas. Programa de Pós-Graduação em Direito. Maceió, 2023.

Bibliografia: f. 127-137.
Apêndices: f. 138-161.

1. Proteção de dados pessoais. 2. Decisão judicial - Explicação. 3. Inteligência artificial. 4. Decisão judicial - Motivação. I. Título.

CDU: 343.451

IZAAC DUARTE DE ALENCAR

"A Proteção de dados pessoais e Inteligência Artificial: tratamento automatizado e o direito à explicação"

Dissertação apresentada ao Programa de Pós-Graduação em Direito da Faculdade de Direito de Alagoas – UFAL, como requisito parcial à obtenção do grau de Mestre.

Orientadora: Prof.ª Dr.ª Juliana de Oliveira Jota Dantas

A Banca Examinadora, composta pelos professores abaixo, sob a presidência do primeiro, submeteu o candidato à defesa, em nível de Mestrado, e o julgou nos seguintes termos:

Prof. Dr. **Marcos Augusto de Albuquerque Ehrhardt Júnior** (UFAL)

Julgamento: APROVADO Assinatura: MARCOS AUGUSTO DE ALBUQUERQUE EHRHARDT J Assinado de forma digital por MARCOS AUGUSTO DE ALBUQUERQUE EHRHARDT J Dados: 2023.11.22 17:06:54 -03'00'

Prof. Dr. **José Barros Correia Júnior** (UFAL)

Julgamento: APROVADO Assinatura: Professor José Barros Correia Junior Assinado de forma digital por Professor José Barros Correia Junior Dados: 2023.11.22 17:29:12 -03'00'

Prof. Dr. **Leonardo Carneiro da Cunha** (UFPE)

Julgamento: APROVADO Assinatura: LEONARDO JOSE RIBEIRO COUTINHO BERARDO CARNEIRO DA C:88769640459 Assinado de forma digital por LEONARDO JOSE RIBEIRO COUTINHO BERARDO CARNEIRO DA C:88769640459 Dados: 2023.11.22 18:19:26 -03'00'

Maceió-AL, 22 de novembro de 2023.

Aos meus amados e saudosos pais, Izaac e Elizabeth (*in memorian*), pois não seria quem sou sem eles.
À minha irmã, Luciene, que sempre me auxiliou quando foi preciso.
À minha eterna namorada e amada esposa Ceres, por sempre estar ao meu lado em todos os momentos.

AGRADECIMENTOS

Agradeço, primeiramente, à Deus, por iluminar meu caminho durante esse jornada. Embora nem sempre claro o caminho traçado, tenho absoluta certeza que sempre esteve ao meu lado.

Agradeço sempre aos meus pais, Izaac e Elizabeth (*in memorian*), por se sacrificarem com intuito de me proporcionar uma formação de qualidade. Sem este apoio, nunca teria chegado até aqui.

Agradeço à minha amada esposa Ceres Louise pelo apoio incondicional, carinho, paciência e por sempre acreditar em minha capacidade.

Agradeço à estimada e sempre gentil orientadora, Profª. Drª. Juliana de Oliveira Jota Dantas, por todos os conselhos, suas orientações, pela paciência e ajuda nesse período.

Agradeço aos professores que compuseram a banca avaliadora, por aceitarem o convite e analisaram esta dissertação.

Agradeço aos amigos e colegas da Pós-Graduação em Direito Público, pelo apoio mútuo e consideração, pois dividimos inúmeras vezes as expectativas, ansiedade, frustrações e sucessos.

Agradeço, igualmente, aos professores que fizeram parte dessa jornada, dividindo seu conhecimento e experiência, em especial aos estimados professores Marcos EhrHardt Jr. e José Barros Correia Júnior por suas orientações e pelas oportunidades.

Quero agradecer a UFAL e seus servidores, principalmente os secretários do PPGDP, que esclareceram as dúvidas e auxiliaram durante todo o curso.

“Existe apenas uma luz na ciência,
e acendê-la em qualquer lugar é iluminar todos os lugares”
Asimov, Isaac

RESUMO

O uso de sistemas de Inteligência Artificial é comum em muitos países, incluindo o Brasil, abrangendo diversas aplicações, inclusive no campo jurídico, como o apoio à tomada de decisões judiciais. No entanto, mesmo quando a decisão é gerada com o auxílio da Inteligência Artificial, a obrigação de motivar essa decisão ainda persiste. Como os processos judiciais envolvem dados pessoais, a Lei Geral de Proteção de Dados (Lei nº 13.709 de 2018) se aplica, garantindo que as partes, como titulares desses dados, tenham o direito de receber uma explicação para essas decisões automatizadas. No entanto, os modelos de Inteligência Artificial usados para auxiliar na tomada de decisões geralmente são opacos, o que torna difícil compreender seu funcionamento, podendo prejudicar as partes cerceando e violando direitos. O principal objetivo deste trabalho consiste em identificar o contorno para a concretização do direito à explicação de decisões automatizadas aplicado à motivação das decisões judiciais realizadas com auxílio da Inteligência Artificial. Para isso, foram realizadas análises de documentos normativos e técnicos pertinentes ao problema de pesquisa, cujos dados resultantes desse procedimento foram confrontados e avaliados com os resultados provenientes da literatura. Assim, observa-se que a transparência e a informação significativa formam os pilares de uma explicação efetiva, da qual o titular de dados, como parte do processo, poderá se valer para compreender uma decisão criada com auxílio da automatização, certificando se houvera ou não violação de direitos. Outro fator importante está na publicização dos algoritmos, necessária para a manutenção da lisura processual, quanto a Inteligência Artificial é utilizada, pois os sistemas inteligentes não são acessíveis em seu código fonte. Por isso, a impossibilidade de explicação do tratamento de dados por Inteligência Artificial, por conta da opacidade desses tipos de sistemas, tem como consequência a inviabilidade do exercício pleno do direito à explicação das decisão automatizadas chanceladas pelo juízo, comprometendo a concretização da motivação das decisões judiciais.

Palavras-chaves: Proteção de dados pessoais, Direito à explicação de decisões automatizadas, Inteligência Artificial, Motivação de decisões judiciais.

ABSTRACT

The use of Artificial Intelligence systems is common in many countries, including Brazil, encompassing various applications, including the legal field, such as supporting judicial decision-making. However, even when the decision is generated with the assistance of Artificial Intelligence, the obligation to justify this decision still exists. Since judicial processes involve personal data, the General Data Protection Law (Law No. 13,709 of 2018) applies, ensuring that parties, as data subjects, have the right to receive an explanation for these automated decisions. However, the Artificial Intelligence models used to assist in decision-making are often opaque, making it difficult to understand their functioning, potentially harming the parties by restricting and violating rights. The main objective of this work is to identify the framework for the realization of the right to an explanation of automated decisions applied to the motivation of judicial decisions made with the assistance of Artificial Intelligence. To achieve this, analyses of normative and technical documents relevant to the research problem were carried out, and the resulting data from this procedure were compared and evaluated with the results from the literature. Thus, it is observed that transparency and meaningful information form the pillars of an effective explanation, from which data subjects, as part of the process, can rely on to understand a decision created with automation's assistance, ensuring whether or not there has been a violation of rights. Another important factor is the disclosure of algorithms, necessary for the maintenance of procedural fairness when Artificial Intelligence is used, as intelligent systems are not accessible in their source code. Therefore, the impossibility of explaining the treatment of data by Artificial Intelligence, due to the opacity of these systems, results in the impracticability of exercising the full right to an explanation of automated decisions endorsed by the judiciary, compromising the realization of the motivation of judicial decisions.

Keywords: Personal data protection, Right to explanation of automated decisions, Artificial Intelligence, Motivation of judicial decisions.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A proteção de dados pessoais engloba múltiplos tipos de proteção e uma delas é a proteção jurídica. No Brasil, a lei nº 13.709/2018, conhecida como Lei Geral de Proteção de Dados, trouxe uma proteção jurídica maior em relação ao tratamento de dados pessoais, empoderando os titulares dos dados, atribuindo-lhes direitos e limitando o uso deste tipo específico de dados por terceiros. De tamanha importância, a proteção de dados pessoais fora absorvida pela Constituição Federal como um direito fundamental, constando no artigo 5º, inciso LXXIX. Tais dados são utilizados para as mais diversas finalidades, dentre elas estão: a criação de perfis de crédito, de seguro, de relações trabalhistas, entre outras. Diante do considerável volume de dados disponíveis atualmente, tornou-se natural a utilização de programas de computadores na diminuição do esforço humano para o devido tratamento destes dados.

A tecnologia computacional avança com o passar do tempo, aumentando sua capacidade de processamento, o que possibilita a implementação de algoritmos complexos que modelam um comportamento inteligente e evolucionário - Inteligência Artificial. Isto permite que problemas complexos possam ser resolvidos com mais eficiência e eficácia (No contexto da computação, a eficiência é um critério de otimização de método e a eficácia um critério de otimização de tempo) em comparação com o esforço humano ou por implementação de algoritmos tradicionais. A partir da observação deste potencial, modelos de negócios adotam software que utilizam Inteligência Artificial para realizar o tratamento dos mais diversos tipos de dados, incluindo-se os dados pessoais. Esta automação pode representar um risco significativo aos titulares de dados pessoais, impactando-os diretamente em suas vidas, pois, em muitos casos, existe uma tomada de decisão ou um perfilamento executado somente de forma automatizada sem nenhuma interação humana significativa.

Na literatura especializada, pode-se identificar a existência de diversas tecnologias que são classificadas como Inteligência Artificial. Em geral, tais tecnologias se enquadram em duas categorias principais: tecnologias baseadas em regras e tecnologias baseadas em aprendizagem de máquina (KINGSTON, 2017). Nas tecnologias baseadas em regras: 1) se faz necessária uma fonte de regras, que pode ser proveniente da modelagem de conhecimento de especialistas ou documentos contendo políticas ou regulamentos; 2) as regras são programadas no sistema e; 3) o sistema de Inteligência Artificial pode facilmente dar explicações sobre seu processamento, se necessário, descrevendo as regras que foram acionadas e as informações que os acionaram, assim como suas conclusões. Nas tecnologias baseadas em aprendizado de máquina: 1) é necessária uma fonte de dados históricos em larga escala. Alguns desses dados são usados como o conjunto de treinamento do qual as associações são aprendidas e outros dados para validação; 2) uma vez que os parâmetros necessários foram definidos corretamente para o aprendizado ideal, o sistema aprende associações por si e nenhuma programação adicional é necessária; 3) assim, em diversos cenários, o sistema de Inteligência Artificial não pode explicar seu raciocínio além de fornecer informações sobre correlação estatística dos dados.

Visando proteger o titular dos dados pessoais deste tipo de tratamento automatizado e desumanizado, o legislador ordinário incluiu no arcabouço da lei supracitada, o direito à revisão de decisões automatizadas (artigo 20 da Lei Geral de Proteção de Dados), o qual diz que o indivíduo titular dos dados possui o direito de requisitar a análise e revisão de decisões que foram tomadas exclusivamente por meio de processos automatizados de tratamento de dados pessoais. O exercício deste direito deve ser acompanhado por uma explicação significativa do processo automatizado de tratamento, pois a oposição a determinada decisão pode ser fundamentada na compreensão de como ela foi tomada. Esta explicação é possível por intermédio do exercício do direito à explicação, contemplada no § 1º do mesmo dispositivo normativo citado, do qual diz que o responsável pelo controle dos dados deve fornecer informações claras e adequadas, sempre que solicitado, sobre os critérios e os procedimentos utilizados para a tomada de decisão automatizada, respeitando os segredos comerciais e industriais.

A apresentação da explicação citada no parágrafo anterior visa substanciar o titular dos dados para que este possa decidir sobre sua oposição a uma determinada decisão automatizada ou enquadramento a um determinado grupo (perfilamento). Esta obrigação imputada ao controlador é proveniente do princípio da transparência (artigo 6º, VI) e não há garantias que sua implementação forneça a segurança adequada da qual o titular dos dados efetivamente necessita.

No mesmo sentido, a Regulamento Geral de Proteção de Dados (*General Data Protection Regulation*), do qual inspirou a lei brasileira, também disciplina o tratamento de dados pessoais por programas totalmente automatizados, o que se inclui tratamento executado por Inteligência Artificial. Neste regramento, há a previsão de intervenção humana para garantia de que as decisões ou perfilamentos não violem quaisquer direitos do titular de dados (EU, 2016). Tal previsão não se encontra na lei brasileira, tornando-a inicialmente facultativa aos controladores.

Esse direito à explicação também se aplica na esfera do Poder Judiciário, no escopo do processo judicial, especialmente quando se trata de decisões judiciais criadas com o auxílio da tecnologia de Inteligência Artificial. Isso por conta da necessidade em que o juízo tem de motivar suas decisões, pois o devido processo legal exige tal justificativa em decidir. Dessa forma, surge a questão de saber se o direito à explicação de decisões automatizadas pode ser empregado como meio de concretizar o princípio da motivação de decisões judiciais, proporcionando proteção ao titular de dados como parte do processo.

No processo judicial, a utilização de sistemas de inteligência artificial na tomada de decisão judicial pode ter consequências negativas significativas, uma vez que a justificação para a decisão tomada pode ser difícil ou impossível de ser compreendida. Essa falta de transparência pode levar a decisões judiciais injustas e discriminatórias. Por exemplo, um sistema com Inteligência Artificial pode aprender a tomar decisões com base em características irrelevantes para o processo, o que violaria o devido processo legal. Se os motivos por trás dessas decisões não forem explicados adequadamente, as partes envolvidas podem não ter a chance de contestar a decisão de maneira também adequada.

Neste contexto, o principal objetivo investigar como a aplicação do direito à explicação de decisões automatizadas, trazido pela Lei Geral de Proteção de Dados, pode concretizar o princípio da motivação de decisões judiciais, quando estas são realizadas com auxílio de sistemas com Inteligência Artificial. Para este fim, pretende-se investigar as limitações da tecnologia em estudo e sua relação com os direitos e garantias fundamentais. Objetiva-se, igualmente, analisar a conexão entre a proteção de dados pessoas e o uso de sistemas de Inteligência Artificial no ordenamento jurídico brasileiro. Ainda assim, intenta-se examinar a forma de explicação de decisões automatizadas, no contexto de Inteligência Artificial, adequada para cumprimento do estabelecido em lei. Por fim, averiguar a suficiência do instrumento da explicação aplicada no âmbito do processo judicial.

Para isso, a metodologia de pesquisa mostra como a pesquisa será conduzida, isto é: a forma pela qual a revisão de literatura será realizada, como se dará a extração de dados relevantes das referências incluídas, a maneira como serão analisados e sistematizadas tais dados, assim como sua exposição. Outrossim, como será realizada as análises normativas pertinentes por meio da hermenêutica jurídica e da releitura do Direito através de uma perspectiva tecnológica, principalmente pela tecnologia da informação.

Ao se considerar diretrizes definidas por GIL 2002 e MOREIRA 2005, a metodologia que foi aplicada nesta pesquisa pode ser classificada como descritiva e exploratória. Em uma análise descritiva, o objetivo principal consiste em descrever certo fenômeno ou população. Por outro lado, em uma pesquisa exploratória, o principal objetivo reside na ampliação do conhecimento em determinado tema. Quanto ao procedimento, o projeto pode ser classificado como pesquisa bibliográfica e documental. Em relação à abordagem utilizada, se trata de uma pesquisa jurídica-tecnológica, pois se faz necessário trazer o aspecto interdisciplinar para a discussão do tema proposto.

Nesta pesquisa, em razão da interdisciplinaridade do tema, foi utilizada a metodologia dedutiva de revisão bibliográfica, com as etapas distintas: 1) a revisão de literatura sobre os temas que são foco desta pesquisa (direito à explicação de decisões e perfilamento automatizados, inteligência artificial e proteção de dados pessoais), da qual tem como propósito fornecer os fundamentos necessários para a compreensão do cerne desta pesquisa (os protocolos de revisão se encontram nos Apêndices A e B); 2) revisão e análise normativa sobre os principais instrumentos normativos nacionais sobre proteção de dados que podem impactar o direito à explicação de decisões automatizadas, cujo objetivo reside no entendimento normativo destinado à regular o objeto da pesquisa; 3) testar as hipóteses levantadas através do método de validação lógica com base na literatura levantada e dos instrumentos normativos analisados e; 4) após as fases citadas, ocorrerá a escrita da dissertação.

Este trabalho tem como abrangência o direito à explicação de decisões automatizadas no contexto da proteção de dados pessoais e tratamento automatizado realizado por Inteligência Artificial, dentro do sistema jurídico normativo brasileiro, após a entrada em vigor da Lei Geral de Proteção de Dados Pessoais. Outras formas de tratamento de dados realizados serão

abordados apenas como exemplificação, não sendo objeto de estudo da presente pesquisa. O mesmo para legislação estrangeira, pois o foco reside no ordenamento jurídico brasileiro.

Como análise de aplicação do direito à explicação citado no parágrafo anterior, somente serão consideradas as decisões judiciais que são geradas com auxílio da tecnologia de Inteligência Artificial, restando fora do escopo as demais atividades realizadas pelo Poder Judiciário com o uso da tecnologia apontada.

Quanto ao aspecto do processo judicial, a finalidade do estudo está na concretização do princípio da motivação de decisões judiciais. Portanto, aos demais princípios, embora relevantes, estão fora do contexto da pesquisa, podendo ser citados, mas não serão discutidos com profundidade, como por exemplo a Inteligência Artificial e provas dentro do processo.

Esta seção descreve a estrutura deste documento, isto é, como ele foi construído. O Capítulo 2 apresenta, com base na tecnologia, a Inteligência Artificial, o ambiente em que está inserida a tecnologia e a relação dos direitos e garantias fundamentais neste ambiente. No Capítulo B, discute-se a proteção de dados pessoais e sua relação com a Inteligência Artificial, assim como seu impactos nos direitos e garantias fundamentais, a regulamentação da tecnologia no Brasil, e as consequências da não observação da legislação para o tratamento de dados automatizado. No Capítulo 4, se expõe as características de uma explicação suficiente desses tratamentos automatizados com base nos níveis de transparência e na informação significativa. Já no Capítulo 5, é discutido a concretização do princípio da motivação das decisões judiciais, quando estas são realizadas com auxílio de sistemas de Inteligência Artificial, através da aplicação do direito à explicação de decisões automatizadas, direito trazido pela Lei Geral de Proteção de Dados.

Embora a Lei Geral de Proteção de Dados esteja focada especificamente no “direito à proteção de dados pessoais”, ela preocupa-se com os efeitos da tomada de decisão automatizada sobre os direitos e liberdades fundamentais das pessoas naturais, incluindo as consequências advindas de possível discriminação por programação ou de uma decisão judicial injusta. Embora existam pesquisas em desenvolvimento para propor soluções jurídicas a respeito do tratamento de dados pessoais por sistemas de Inteligência Artificial, não há, contudo, uma pacificação para uma solução viável, restando os titular dados incerteza de uma tutela jurisdicional efetiva.

2 DA RELAÇÃO ENTRE AMBIENTE DIGITAL, INTELIGÊNCIA ARTIFICIAL E DIREITOS FUNDAMENTAIS

O propósito principal deste capítulo é apresentar uma base teórica essencial para compreender a relação entre dados e sistemas de Inteligência Artificial, abordando conceitos comumente empregados na literatura especializada, técnicas e aplicações relevantes, bem como destacando a importância e as consequências para o desenvolvimento da sociedade. Adicionalmente, este capítulo busca examinar a interação entre o ambiente digital, Inteligência Artificial e dados com os direitos e garantias fundamentais, avaliando brevemente os impactos gerados por essas tecnologias.

Para esta finalidade, primeiramente é preciso entender o ambiente digital, com suas limitações, e seu relacionamento com o ambiente analógico, assim como compreender a importância dos dados para esse ambiente existir. Após, se expõe os reflexos dos direitos e garantias fundamentais nesse espaço digital, pois o direito o alcança e, por essa razão, entender essa conexão é de vital relevância para o estudo deste capítulo. Em seguida, conceitos e apontamentos sobre software e algoritmo também são relevantes, uma vez que através deles as ações humanos são realizadas e representadas dentro do ambiente digital e entender suas limitações se torna crucial para discussões futuras.

Por fim, as bases fundamentais para entender minimamente a Inteligência Artificial é mostrada, abarcando alguns dos seu conceitos mais aceitos na literatura e a relação intrínseca com a autonomia tecnológica através da aprendizagem de máquina. Todos os conceitos trazidos, fornecem a um entendimento mais apurado do funcionamento da tecnologia, permitindo uma análise com mais acurácia de temas do direito, como responsabilidade civil e o objetivo deste trabalho, por exemplo.

Isto se faz necessário, pois, o objeto de pesquisa envolve decisões automatizadas por sistemas inteligentes. Assim sendo, buscou-se por publicações científicas, teses, dissertações, monografias, periódicos científicos, livros e outras produções, incluindo-se as técnicas que ajudaram no corte necessário para o avanço da análise e discussão nos próximos capítulos. Foi realizada uma revisão de literatura narrativa, contando com o registro do protocolo simplificado, no Apêndice B.

2.1 Breves explanações sobre o ambiente digital e seu relacionamento com dados

O desenvolvimento do ambiente digital se confunde com o desenvolvimento da computação digital. Ambos precisam necessariamente de uma construção física de interação e processamento no ambiente físico, aqui referido como ambiente analógico. Neste ambiente, é possível lidar com valores infinitos dentro de um determinado intervalo. De fato, o ambiente analógico está sob as limitações das leis da física e da natureza e o ser humano agrega construindo realidades fictícias que interagem diretamente com esse ambiente.

Todas as atividades humanas estão no ambiente analógico ou a partir dele, assim como

seus registros. A principal relação entre esse ambiente e os dados reside na forma de sua persistência, processamento e transmissão. O desenvolvimento de tecnologias como a escrita, assim como diversos meios físicos para sua persistência impulsionaram a própria evolução humana e social. Ainda hoje, tecnologias analógicas são utilizadas, não mais com o mesmo protagonismo, porém mantendo sua utilidade.

Com o advento da computação digital[1], uma variedade significativa de atividades humanas migraram para o ambiente digital, visto que o nível de eficiência e eficácia extraídos em um espectro relevante de aplicações indicam que as máquinas são mais efetivas do que o ser humano em diversos cenários. Este ambiente artificial, criado com base em tecnologias analógicas, utiliza sinais digitais constituídos de zeros e uns, isto é, uma linguagem binária, como cerne de seu funcionamento. Assim sendo, tudo o que adentra a este ambiente digital, para ser reconhecido e processado, precisará estar neste formado, caso o contrário não será compreendido por sistemas digitais.

Já a natureza do ambiente digital é limitada, pois assim como o sinal digital, o ambiente lida com valores discretos[2]. Não há como representar os infinitos valores contínuos[3] encontrados no ambiente analógico. Isto torna tudo o que está no digital igualmente limitado em nível de processamento e poder computacional, incluindo os softwares de computador.

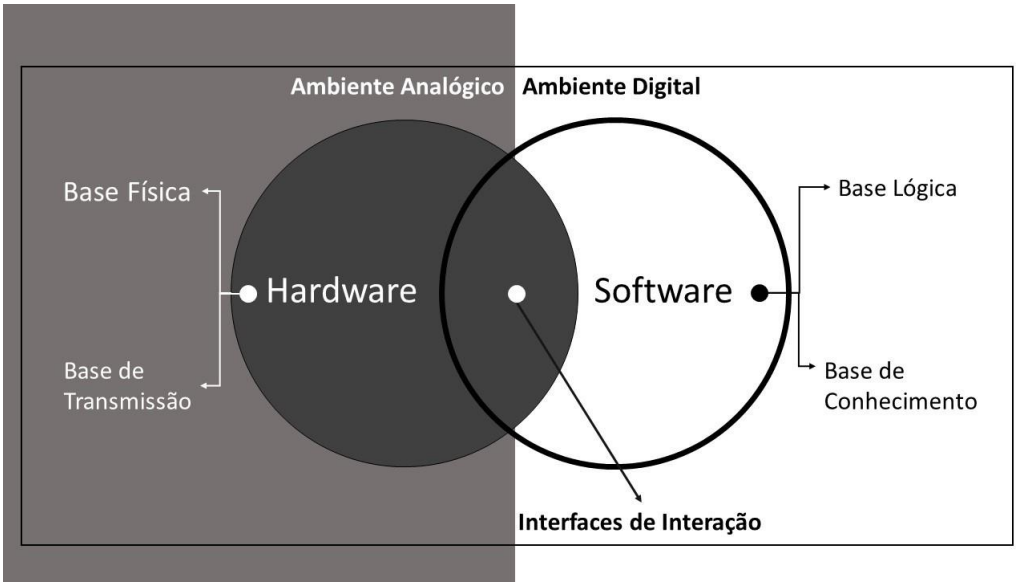
A ISO (2015) define o ciberespaço como "ambiente complexo resultante da interação de pessoas, software e serviços na Internet por meio de dispositivos de tecnologia e redes conectadas a ele, que não existe em qualquer forma física". A definição de ambiente digital, no que concerne ao ciberespaço, consistem em um completo sistema de hardwares de softwares funcionando em conjunto, um dependendo do outro para atingir seu propósito. Desta forma, pode-se apontar que o ambiente digital é uma abstração simbólica do ambiente analógico. Essa interação pode ser expressa a partir do diagrama de *Venn* contido na Figura 1:

[1]Computação digital é o estudo dos sistemas computacionais digitais, que são dispositivos eletrônicos que manipulam informações digitais utilizando sequências de bits. Esses sistemas são baseados em circuitos lógicos que podem ser programados para executar diversas tarefas, desde operações matemáticas simples até a execução de programas complexos. A computação digital é a base para a tecnologia da informação e é fundamental para o desenvolvimento de sistemas computacionais modernos (PATTERSON; HENNESSY, 2016).

[2]Valores discretos são valores que representam uma contagem, número de itens ou uma quantidade inteira, geralmente representados por números inteiros. Eles diferem dos valores contínuos, que representam uma grandeza infinita e são representados por números reais (LAROSE; LAROSE, 2014).

[3]Um valor contínuo é um tipo de dado numérico que pode assumir uma infinidade de valores dentro de um intervalo contínuo. O conjunto de valores possíveis é ilimitado e inclui, em geral, todos os números reais (LARSON; FARBER, 2019).

Figura 1 – Ambiente Digital.



Fonte: o autor.

A Figura 1 mostra a integração entre os ambiente analógico e o ambiente digital. Nele, é possível visualizar uma área de interseção entre os ambientes, da qual ambos se confundem. Percebe-se que existe uma simbiose entre os ambientes, pois os elementos consistentes no ambiente analógico, os hardwares, não detém funcionalidade sem os elementos do ambiente digital, os softwares e estes, por sua vez, também necessitam do hardware para sua existência.

Como elemento primordial representado no ambiente analógico está o hardware, do qual agrega as bases físicas e de transmissão. Esta junção ocorre por afinidade de funções. A base física consiste em elementos que contêm o processamento, a exemplo dos computadores, smartphones, tablets, entre diversos outros. Na base de transmissão, encontram-se os elementos de infraestrutura de comunicação e transferência de arquivos e dados, como exemplo cita-se: o cabeamento de rede e os elementos de comutação de conexão (hubs, switchs e roteadores).

Para o funcionamento do ambiente digital é necessário a construção de um conglomerado complexo de infraestrutura conectada entre si e capaz de prover os recursos requeridos para executar e manutenção do ambiente. Portanto, a existência de um software funcionando em uma máquina isolada e sem conexão externa é possível, porém ineficaz e ineficiente nos moldes da atual sociedade hiperconectada, salvo em necessidade estrita de controle de ambiente. Por esta razão, o alcance de ambiente digital se estende por todos os continentes através de meios de transmissão diversos, desde cabos submarinos até elos por satélites.

Por outro lado, no ambiente digital, é possível apontar como elemento principal, o software. Este é, igualmente, dividido em duas partes: a base lógica e a base de conhecimento. A primeira, pode-se listar uma séria de categorias de software também por afinidade de funcionamento, a exemplo dos sistemas operacionais, apps para mobile, plataformas web, softwares para desktop, além de todos os protocolos de comunicação utilizados em redes de comunicação. Para a base de conhecimento, se tem os dados e as informações.

Por fim, existe uma área de interseção entre os ambientes, onde se pode encontrar as interfaces de interação humano-máquina e máquina-máquina. Entende-se como interface de interação todos os dispositivos e softwares que servem como ponte para comunicação através de *Inputs* e *Outputs*. No diagrama, é possível perceber certo nível de abstração entre os ambientes, pois pode-se questionar o porquê de sua construção com uma área de interseção e não com o conjunto que representa o ambiente digital contido no ambiente analógico devido a dependência do físico para existência do digital. Todavia, a computação evoluiu de tal forma, que atualmente não é preciso ter determinado software instalado no próprio dispositivo para utilizá-lo, precisando apenas do acesso à Internet. Desta forma, é preferível representar este relacionamento com certo grau de liberdade em termos de abstração.

Com a invenção do transistor, em 1948, nos Laboratórios Telefônicos da Bell, foi possível substituir as válvulas eletrônicas nos computadores digitais, o que simplificou o processo de indicação de tensão e não tensão, representando uns e zeros, consequentemente. Naquele mesmo ano, Claude Shannon apresentou o conceito de bit, formulando o conceito de codificação de dados mediante um átomo binário. Este conjunto sustentou a base do processamento e da computação moderna. Quando se alinha, portanto, 8 bits em conjunto, se tem os Bytes[4], possibilitando a representação de letras ou números, isto é, um dado (BARBIERI, 2019).

Os dados agora poderiam ser representados no ambiente digital e não somente no ambiente analógico. No entanto, o fundamento se mantém o mesmo. Dados são símbolos, cujo valor semântico, ou seja, seu significado possui baixa relevância para compreensão e entendimento de algo. Para Setzer (1999), o dado é como uma sequência de símbolos quantificáveis ou quantificados. Barreto (2005) define dados brutos como simples representação de fatos, texto, gráficos, sinais e que, depois de processados resulta em informação. Já para Semidão (2014), um dado se configura como elemento nuclear e elementar, como um fato ou átomo, desprovido de significado imediato, sendo algo que é perceptivo, porém de baixo teor semântico. Portanto, o dado precisa ser tratado para enriquecer seu significado e, a partir disto, proporcionar maior nível de compreensão e interpretação.

Com as definições de dado trazidas no parágrafo anterior, vislumbra-se que são entidades matemáticas, representados formalmente. Agregando como características a quantificação, podem ser armazenados em um dispositivo computacional e por consequência são processáveis por ele. Portanto, dados e informação não são a mesma coisa, nem tem o mesmo conceito, sendo o segundo resultado do tratamento do primeiro.

Os dados são classificados em diversas categorias, inclusive por sua forma de persistência. As formas de persistências de dados em ambiente digital, são elas: 1 - forma não estruturada; 2 - forma estruturada e; 3 - forma semiestruturada. A persistência não estruturada

[4]Um byte é uma unidade de medida de armazenamento de dados que consiste em 8 bits de informação. Ele é a unidade fundamental da memória de um computador e é usado para representar um único caractere alfanumérico (HENNESSY; PATTERSON, 2014).

funciona agregando dados dentro de arquivos na estrutura de diretório e sistemas de arquivos do sistema operacional e como não há uma estrutura que facilite seu manuseio exige esforço de programação considerável, isto porque o próprio software, além de todas as funcionalidades para que foi destinado ainda teria que lidar com os dados. Não existe, portanto, uma estrutura fixa e rígida a partir de definições prévia, assim como exitem em dados que são estruturados.

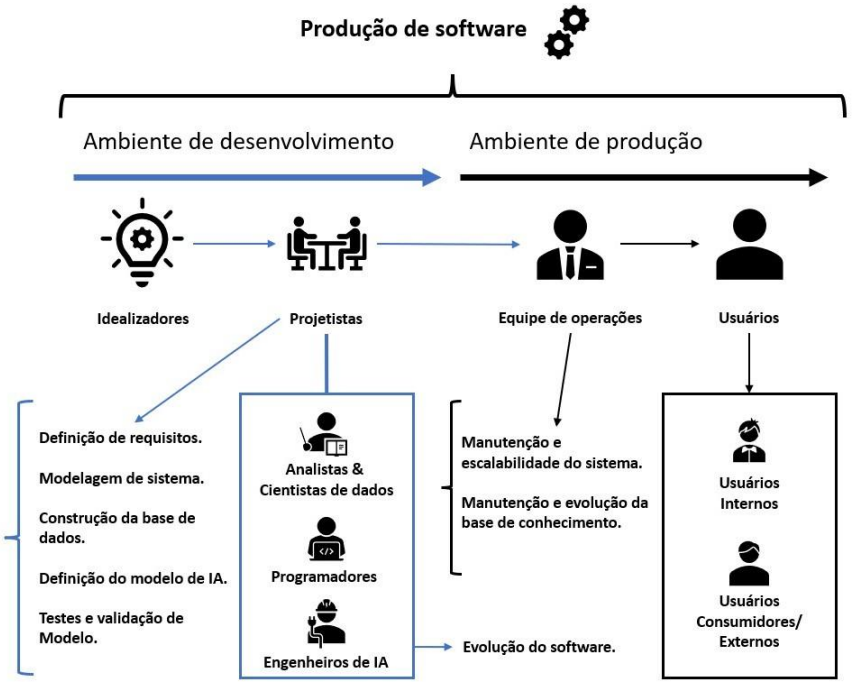
Por sua vez, os dados estruturados são organizados em blocos semânticos modelados e planejados antes mesmo da inserção dos dados. Exemplos usualmente associados são os bancos de dados relacionais construídos como estruturas em linhas e colunas. Os Sistemas de Gerenciamento de Banco de Dados (SGBD)[5] são capazes de fornecer todas as ferramentas necessárias para as atividades de gerenciamento e gestão de dados para o suporte das atividades das instituições. Já os dados semiestruturados agrupam características dos dados estruturados e não estruturados simultaneamente, oferecendo uma estrutura flexível diferente das formalidades impostas aos dados estruturados e da total falta de estrutura dos dados não estruturados.

2.2 Apontamentos sobre algoritmos e softwares

Os softwares representam a base lógica para a existência do ambiente digital, como mostrado na Seção 2.1, constituindo o meio pelo qual se concretizará os direitos e garantias fundamentais dentro desse ambiente, como evidenciado na Seção 2.4. Por conta disto, ele pode impactar positivamente ou negativamente o indivíduo ou a sociedade, sendo necessária que, em sua concepção, já exista a preocupação em observar tais direitos e garantias. Na Figura 2 é possível visualizar as etapas do construção do software em conjunto com os recursos humanos associados.

[5]Um sistema de gerenciamento de banco de dados é projetado para auxiliar a manipulação de vastos conjuntos de dados através de várias ferramentas que o compõe (RAMAKRISHNAN; GEHRKE, 2008).

Figura 2 – Cadeia de existência do software de computador.



Fonte: o autor.

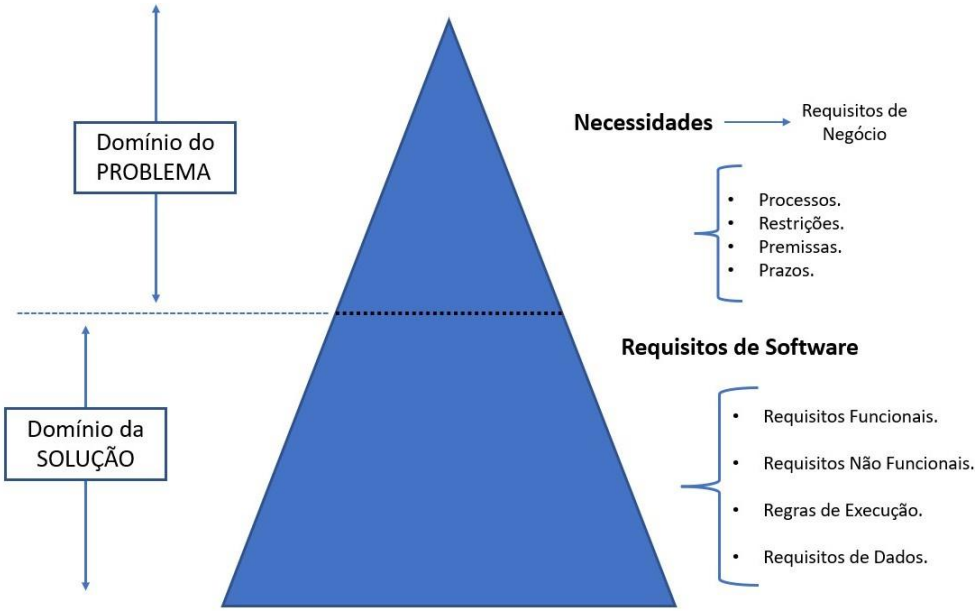
Para a confecção de um artefato de software há duas etapas fundamentais: a fase de desenvolvimento e a fase de produção. Na primeira, o software será planejado, implementado, testado e validado para que, após, possa ser utilizado para o que foi criado. No planejamento do software, os requisitos são levantados e, a partir deles, os algoritmos são criados. Na implementação, os algoritmos são traduzidos para alguma linguagem de programação, escolhida na etapa de planejamento, e convertidos para linguagem de máquina, criando o software propriamente dito. Com o software preparado, os testes são realizados e, com resultados satisfatórios, validado. Contudo, salienta-se que não há a obrigatoriedade deste processo acontecer em cascata, esperando-se que a etapa anterior finalize para o começo da próxima. O processo, a depender da gestão, pode seguir de forma concorrente.

Na segunda, o software encontra-se produzindo resultados para a atividade ao qual se destina. Este funcionamento exige um comportamento eficiente e eficaz, sempre dentro do contexto do qual está inserido. Se na primeira etapa os projetistas são os protagonistas, na segunda a equipe de operações entra em cena com a responsabilidade de monitorar o sistema a fim de conduzir o funcionamento do sistema de maneira satisfatória. A manutenção do software, assim como sua evolução - remetendo-se a atualização em novas versões - podem ocorrer periodicamente a partir desta segunda etapa. Portanto, apesar de existir uma nítida separação de responsabilidades, ocorre constante interação entre projetistas e operadores a fim de melhorar o sistema como um todo, pois ajustes são esperados no decorrer da vida útil do software.

Remetendo-se para a etapa de desenvolvimento, os requisitos necessários para construção da solução de software são levantados por analistas de sistemas, cientistas da

computação e engenheiros de software que fazem a interface entre a atividade de negócio e os implementadores que automatizarão tal atividade. Há, em um projeto de desenvolvimento, uma variedade de tipos de requisitos, porém dois grupos se destacam: a) os requisitos de negócio, que descrevem as necessidades de negócio e da atividade dos idealizadores e; b) os requisitos de software, dos quais remetem as ações em que o software executará de forma a automatizar os processos de negócio. Esta interação pode ser visualizada na Figura 3.

Figura 3 – Requisito para construção de um software.



Fonte: o autor.

Os requisitos de negócio são impostos pelos idealizadores, aqueles que impulsionam o processo de desenvolvimento do software, como contratantes ou a direção de uma instituição que possui sua própria equipe de desenvolvimento. Esses requisitos incorporam as regras de negócio da atividade dos idealizadores e a partir desta imposição que os demais requisitos serão definidos, a exemplo dos requisitos de software. Esses requisito, por sua vez, possuem subdivisões, das quais se ressaltam os requisitos funcionais e requisitos não funcionais. Os primeiros são requisitos que indicam as ações que serão automatizadas e refletirão os processos de negócio ou atividade dos idealizadores, como o gerenciamento de clientes, produtos, serviços e outros. Os segundos são ações de suporte aos primeiros, portanto, os sustentando em execução como recursos, a exemplo da aplicação de criptografia, processamento diferenciado, armazenamento distribuído, dentre outros.

Definidos os requisito, segue-se a modelagem do software, da qual modelos são criados com o fim de explicar as características e comportamento do sistema que será criado. Um modelo pode ser definido como uma simplificação da realidade e seu objetivo consiste na compreensão de fenômenos, fatos ou sistemas complexos. Assim, tais modelos são utilizados para delimitar

o problema de domínio do software, permitindo, inclusive, sua divisão em problemas menores, o que resulta na solução do problema de maior interesse.

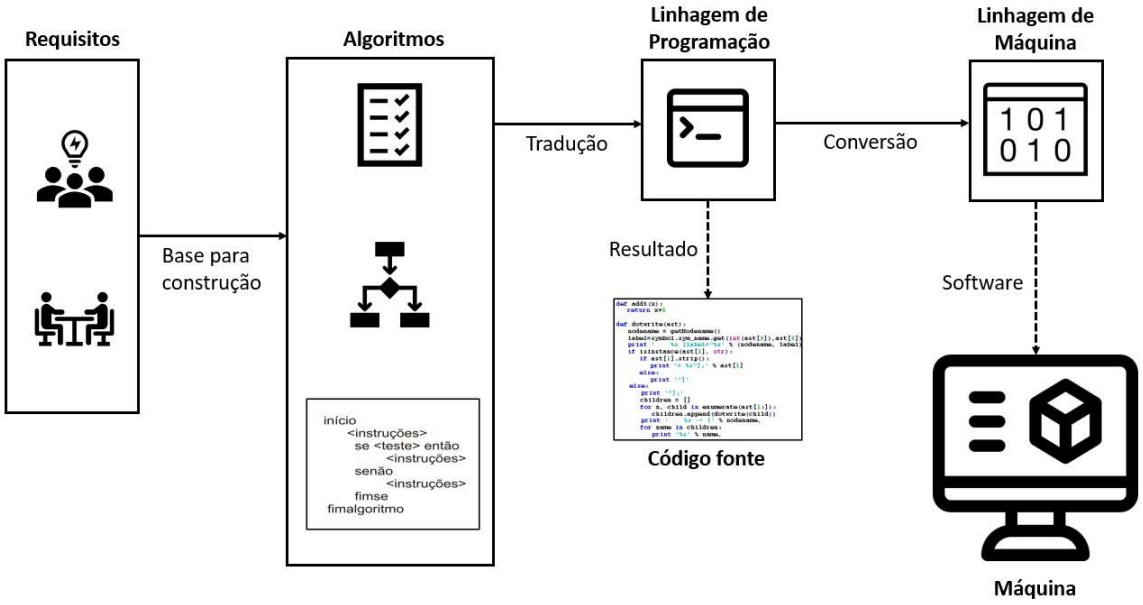
Do entendimento proporcionado pela modelagem, os algoritmos são construídos. Um algoritmo é formado por um conjunto de instruções organizadas, logicamente encadeadas e finitas - dentro de uma estrutura de entrada, processamento e saída - cujo objetivo se destina à resolução de um ou mais problemas específicos. Os algoritmos não estão adstritos à computação, portanto podem ser abstraídos para utilização em diversas áreas de aplicação e apesar de sua execução ser direta ou indireta, sempre será através de uma entidade externa, pois os algoritmos não são autoexecutáveis. Como entidade pode-se citar o ser humano executando diretamente instruções de montagem de um móvel contidos em suporte analógico como papel ou a máquina executando indiretamente o algoritmo através de um software.

Os algoritmos podem ser classificados por diversas perspectivas, são elas: 1) por sua representação, classificação que indica como o algoritmo será apresentado. Essa classificação possui subdivisões com o fim de melhorar o entendimento do próprio algoritmo, são elas: representação narrativa-descritiva, representação gráfica, representação matemática, representação em pseudocódigo e representação híbrida; 2) por sua forma de execução, a qual indica o comportamento do algoritmo quando for executado por entidade externa, possuindo subdivisão: algoritmos determinísticos e algoritmo não-determinísticos; 3) por sua área de aplicação, onde o algoritmo é aplicado de acordo com a área de conhecimento, como algoritmos computacionais ou algoritmos de processos; 4) por sua estrutura, que aponta as técnicas utilizadas para sua construção, possuindo subdivisões: algoritmos construídos com técnicas de tradicionais de programação e algoritmos construídos com técnicas avançadas, como as de Inteligência Artificial e; 5) por sua finalidade, essa classificação agrupa os algoritmos de acordo com seu objetivo de construção: algoritmos de otimização, algoritmos de busca, algoritmos de classificação e muitos outros.

No que se refere aos algoritmos computacionais, classificação referente à área de aplicação, estes precisam necessariamente de tratamento específico para que as máquinas, entidades que os executarão, possam compreendê-los. Para isso, os algoritmos serão traduzidos para uma linguagem de programação, usualmente de alto nível[6], resultando em um código fonte, do qual ainda necessitará de conversão para a linguagem de máquina através de processos específicos, constituindo, assim, o software que será executado pela máquina, isto é, pelo computador digital. Este processo pode ser visualizado na Figura 4.

[6]Linguagens de programação podem ser de alto nível ou de baixo nível. O primeiro diz respeito às linguagens em que o ser humano se comunica, portanto compreensíveis. Linguagens de baixo nível se aproximam do código binário, zeros e uns, possuindo complexidade elevada, pois se faz necessário o conhecimento da arquitetura de hardware exata para programação ao contrário das linguagens de alto nível que abstraem o hardware, não necessitando conhecimento sobre ele para realização da programação.

Figura 4 – Caminho para construção de um artefato de software.



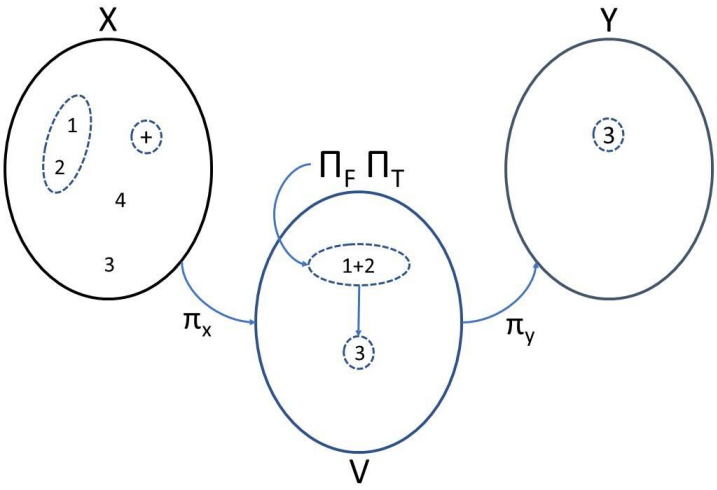
Fonte: o autor.

Por sua vez, o software irá absorver os algoritmos criados, juntando-os com outros elementos que comporão o artefato, como imagens e banco de dados. Um software, como programa de computador, pode ser definido como um conjunto de operações e testes organizados em uma estrutura de controle de fluxo. Em grande parte das vezes, esse software manipulará dados dentro de uma estrutura pré-concebida para melhor eficácia e eficiência dentro de seu contexto de funcionamento.

No entanto, os computadores atuais são limitados e podem ser representados por um modelo constituído por uma 7-tupla, sequência representativa de sete elementos imutáveis, M = {V, X , Y, π_x, π_y, Π_F, Π_T}. Cada elemento do conjunto M (que representa a Máquina) compõe uma funcionalidade para o devido processamento do software dentro de uma arquitetura ao qual todo computador digital clássico moderno é construído. Dentro do modelo, o conjunto V representa os valores que estão na memória volátil principal do computador, usualmente chamada de memória RAM[7]. Esses valores aguardam para serem computados pelo processador/processadores da máquina. Por sua vez, o conjunto X representam os valores de entrada, os quais serão transportados para a memória da máquina e comporão o conjunto V. O conjunto Y representa os valores de saída do processamento, dos quais serão encaminhados para persistência ou visualização posterior. A Figura 5 mostra o modelo de máquina descrito.

[7]Segundo Stallings (2010) , a RAM (*Random Access Memory*) é um componente essencial em qualquer sistema de computação moderno, pois é responsável pelo fornecimento de dados para a CPU (*Central Processing Unit*) de forma rápida e eficiente. Ela permite que os programas sejam executados em tempo real, pois o processador pode acessar a memória de forma aleatória, sem precisar seguir uma ordem sequencial de leitura. A capacidade de armazenamento da RAM é medida em bytes e varia de acordo com o tipo e a configuração do sistema.

Figura 5 – Modelo de máquina.



Fonte: o autor.

Ainda na Figura 5, é possível observar a função π_x, que representa a função de entrada responsável por deslocar valores do conjunto X para o conjunto V, sendo a ordem desses valores determinado pelo tipo de função de entrada escolhida na construção do software. Após o processamento, os resultados são deslocados do conjunto V para o conjunto X pela função de saída π_y. Por fim, há os interpretadores Π_F e Π_T, responsáveis respectivamente pelas operações e testes. Na figura citada, a operação de adição é reconhecida pela máquina, pois já existe sua indicação previamente definida no interpretador de operações Π_F. No que concerne ao interpretador de testes Π_T, embora não esteja representada visualmente na Figura 5, pode-se indicar o uso de estruturas de condição que utilizam a lógica Se-Então-Senão, da qual tem como efeito a mudança de ordem de instruções executadas em tempo real pelo software, modificando seu comportamento.

Essa base conceitual de máquina está sobre quaisquer softwares, portanto, é irrelevante se o software é construído com algoritmos tradicionais ou com técnicas voltadas à Inteligência Artificial. Neste sentido, todo artefato de software deve ser moldado para seu funcionamento com base em entradas, processamento e saídas, levando-se em consideração o que a máquina pode fazer, mesmo que de forma restrita. Isto porque existem limitações em termos de processamento, mesmo que se consiga alcançar resultados vastos e promissores com o nível tecnológico atual.

2.3 Bases fundamentais para o entendimento da Inteligência Artificial

Os sistemas inteligentes tiveram sucesso em sua utilização devido a dois fatores relevantes: a inteligência e o artefato (NORVIG; RUSSELL, 2013). De inicio, cabe definir o que é inteligência para o escopo deste estudo, visto que não há uma pacificação quanto ao seu conceito nas diversas áreas de conhecimento que discutem o tema, pois cada uma delas aborda a inteligência de uma forma particular. O primeiro fator, portanto, pode ser estudado por diversos

prismas: filosófico, matemático, econômico, psicológico, pela perspectiva da neurociência, da engenharia da computacional, etc. O segundo fator constitui o suporte o qual a inteligência funcionará e a máquina digital mostrou-se o meio mais eficiente para a tarefa até então.

Na filosofia, as contribuições trazidas pelo racionalismo auxiliam na compreensão de como o pensamento racional pode conduzir até o conhecimento, que por sua vez, direciona as ações do ser humano. A razão é tida como o núcleo do conhecimento e a mente atribui significado as informações, além de organiza-las e sistematiza-las.

Na tentativa de explicar como o conhecimento é construído através do pensamento, Aristóteles desenvolveu um sistema informal do qual se utiliza de silogismo para gerar conclusões de forma mecânica, com base em premissas estabelecidas anteriormente. Isso permite, de certa forma, que um raciocínio útil possa ser conduzido por um artefato mecânico, como um computador analógico. Apesar da possibilidade de criar um sistema físico que possa utilizar regras lógicas para criar algum grau de conhecimento, não se pode confundir a mente em si com o próprio artefato.

Neste ponto, a discussão sobre a distinção entre mente e matéria trazida pelo dualismo emerge e, consigo, vários problemas são levantados. Dentre eles, a mitigação ou a exclusão do livre-arbítrio tem significativo destaque, pois sistemas físicos são governados por leis da física que se sobressaem a qualquer escolha. Diante desse viés físico, o mesmo ocorre com a máquina, uma vez que suas regras de funcionamento estão pré-programadas, o que elimina, a priori, uma decisão que não esteja pré-estabelecida.

Para o materialismo, entretanto, a preocupação com o livre-arbítrio se reflete apenas em percepções de como as possíveis escolhas se apresentam para uma mente física, restando esta adstrita às leis da física. A mente física manipula o conhecimento de alguma forma e a extrapolação de tal conhecimento ocorrerá com a aquisição de novas experiências através dos sentidos; é o que preconiza o empirismo. Logo, a análise de novas percepções criam novas informações, que serão agrupadas com o conhecimento anteriormente adquirido.

Já no positivismo lógico, há uma junção entre o racionalismo e o empirismo, onde teorias logicamente encadeadas caracterizam o conhecimento. O pensamento racional pode ser construído através das experiências e isto acaba por agrupar características que podem explicar o como da percepção dos sentidos chega-se ao conhecimento, passando primeiramente pelo pensamento racional, implicando na utilização da lógica para alcançar tal resultado.

Apesar do prisma filosófico fornecer um alicerce para o entendimento da inteligência por intermédio do pensamento racional e da criação de conhecimento, foi a matemática que trouxe a escalabilidade para a discussão, visto que marcou a criação de uma ciência formal, da qual exigiu o desenvolvimento de três áreas basilares: a lógica, a computação e a probabilidade. Na lógica, avanços como a criação da lógica proposicional e a lógica booleana impulsionaram a criação da lógica de primeira ordem, até hoje utilizada. Igualmente, foi possível estender a lógica booleana para incluir objetos lógicos, reais e a relações entre eles. Com isso, foi possível inferir logicamente resultados no mundo real.

O próximo passo reside em relacionar a lógica e a computação, delineando as limitações do que se poderia fazer com ambas. Computar consiste em realizar cálculos, ou seja, uma atividade que adentra a seara matemática. Essa atividade se torna mais efetiva quando há a utilização de máquinas, como os atuais computadores digitais. Esse tipo de artefato tem evoluído em sua velocidade e poder de processamento ao longo do tempo, permitindo que mais problemas possam ser resolvidos. Diante disso, a engenharia da computação tem papel fundamental para estabelecimento da Inteligência Artificial. Todavia, se faz necessária a reflexão de que nem todos os problemas conhecidos são tratáveis (resolvidos) devido a elevada complexidade computacional.

Há uma relação inversamente proporcional entre tratabilidade e complexidade computacional, pois a execução de um determinado algoritmo, por intermédio de um software, pode requerer uma quantidade impactante de recursos computacionais para resolução de determinados problemas. Por esta razão, a utilização de sistemas ditos inteligentes exige a moderação de recursos computacionais, devido à otimização que deles se esperam, como uma tentativa de resolver a maior quantidade de problemas com o mínimo de alocação de recursos.

A probabilidade fecha a tríade trazida pela matemática, posto que um sistema puramente lógico não é o mais adequado para lidar com incertezas. A probabilidade, todavia, modela o grau de crença a partir de evidências que estão disponíveis, podendo sofrer alterações em caso de surgimento de novas evidências, incrementando a probabilidade de ocorrência de certo evento. Isso permite recalcular a probabilidade de forma constante, no que se chama de probabilidade dinâmica. Essa constitui parte relevante das abordagens modernas para o raciocínio incerto em sistemas de Inteligência Artificial.

Pelo prisma da matemática, a inteligência pode ser aferida e representada formalmente através da atividade da computação em união com modelos lógicos e probabilísticos. Este último, em especial, lida com sistemas que incorporam a incerteza, inclusive na tomada de decisão, como no campo da economia, cujo objetivo está na realização de escolhas que podem conduzir a resultados preferenciais, através de decisões ótimas. Neste sentido, a teoria da decisão fornece uma estrutura formalizada para decisões sob cenários de incerteza utilizando ferramentas probabilísticas para percepção adequada do contexto da tomada de decisão em si.

A tomada de decisão pode então ser considerada inteligência na medida em que ela direciona ações com base no conhecimento construído pelo pensamento lógico e pela experiência. Por isso, os dados detêm importância impar ao se projetar os sistemas com Inteligência Artificial. Isso porque a partir do tratamento deles haverá a escalada até a inteligência, culminando em ações que interferem na atividade humana de alguma forma.

A inteligência converge, contudo, para um ponto em comum em todos os prismas citados: a ação ou a omissão. O artefato artificial construído para suportar a inteligência, remetendo-se à Seção 2.1, possui limitações em termos de capacidade de processamento, inerente do ambiente digital e dos computadores digitais. Neste sentido, a inteligência também será limitada neste formato, podendo-se realizar quantidade significativa de coisas, mas não tudo.

Essa concepção de Inteligência Artificial teve sua construção iniciada com o intuito de simular o pensamento ou a forma de pensar e agir do ser humano, assim como o de simular o pensar ou o agir racionalmente - direcionando-a novamente ao ser humano. Ambos os aspectos podem incorporar a inteligência pelos diversos prismas, em especial, pelo matemático e computacional, devido a espécie de suporte fornecido pelo artefato artificial. Cabe, então, definir o que é a Inteligência Artificial propriamente dita.

2.3.1 Definições e conceitos de Inteligência Artificial

Conceituar Inteligência Artificial apresenta-se como uma tarefa complexa, assim como conceituar a inteligência natural. Com base no artefato de suporte e no ambiente de implementação, o impulso inicial direciona para uma definição da Inteligência Artificial como um software. Isso significa dizer que sua natureza técnica de existência se relaciona intimamente com a definição de software, apresentada na Seção 2.3. Contudo, apesar de todos os sistemas de Inteligência Artificial serem softwares (atualmente), nem todo software é um sistema de Inteligência Artificial, o que afastaria tal conceito como uma definição suficiente para a tecnologia.

No âmbito jurídico brasileiro, o software é definido no Art. 1º da Lei de nº 9.609 de 1998, conhecida como a lei do software, como sendo ”a expressão de um conjunto organizado de instruções em linguagem natural ou codificada, contida em suporte físico de qualquer natureza, de emprego necessário em máquinas automáticas de tratamento da informação, dispositivos, instrumentos ou equipamentos periféricos, baseados em técnica digital ou análoga, para fazê-los funcionar de modo e para fins determinados”(BRASIL, 1998). A definição legal apresentada está diretamente relacionada à natureza técnica dos sistemas com Inteligência Artificial por conta de seu suporte físico, apenas, o que não é suficiente como conceito da Inteligência Artificial, jurídico ou não.

Uma vez que a representação por software não é suficiente para estabelecer uma definição, cabe direcionar a análise partindo-se dos prismas que lidam com a definição de inteligência de alguma forma. Parte relevante destes prismas se utilizam da inteligência humana como referência para estabelecer um ponto limiar de concepção de uma Inteligência Artificial. O pensamento racional do ser humano é a base para isto, usualmente. O artificial está na tentativa de reprodução do pensamento humano em contexto de simulação. Norvig e Russel agrupam as definições com base no pensamento e no agir humano, como mostra a Figura 6.

Figura 6 – Conceitos de Inteligência Artificial com base no pensamento e agir humano.

Pensando como um humano	**Pensando racionalmente**
"O novo e interessante esforço para fazer os computadores pensarem (...) *máquinas com mentes*, no sentido total e literal." (Haugeland, 1985) "[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado..." (Bellman, 1978)	"O estudo das faculdades mentais pelo uso de modelos computacionais." (Charniak e McDermott, 1985) "O estudo das computações que tornam possível perceber, raciocinar e agir." (Winston, 1992)
Agindo como seres humanos	**Agindo racionalmente**
"A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas." (Kurzweil, 1990) "O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas." (Rich and Knight, 1991)	"Inteligência Computacional é o estudo do projeto de agentes inteligentes." (Poole *et al*., 1998) "AI... está relacionada a um desempenho inteligente de artefatos." (Nilsson, 1998)

Fonte: Norvig e Russell (2013).

Na Figura 6 é possível visualizar o agrupamento de conceitos de Inteligência Artificial pelo viés do pensamento e no agir do ser humano como ser racional. A premissa aplicada aponta para a inserção de inteligência em máquinas para que estas possam executar funções tal como o ser humano as fariam ou até mesmo o superando.

Quando se volta para o pensamento humano, primeiramente se faz necessário entender como o ser humano pensa e determinar se um pulso cognitivo genuíno poderia ser reproduzido artificialmente em uma máquina digital. Isso permite delinear os limites entre o pensamento como fonte da inteligência e pensamento como fonte da autoconsciência. No primeiro, o pensamento poderia ser representado logicamente e formalmente com o propósito de substanciar ações racionais executadas por um artefato artificial, inclusive mecânico. No segundo, o pensamento fornece uma percepção própria de existência. Segundo John Searle, a autoconsciência vai além de métodos lógicos reprodutíveis e é resultado inerente da atividade cerebral através do comportamento de neurônios, fruto de milênios de evolução. Ainda segundo Searle, por mais que se atribua inteligência a uma máquina, ela não seria capaz de alcançar uma consciência propriamente dita (LYRA; MOGRABI; EL-HANI, 2016).

Já na perspectiva do agir como o ser humano ou racionalmente, um sistema com Inteligência Artificial seria indistinguível do ser humano na sua forma de atuar. A máquina absorveria atributos e habilidades que a permitiria ser confundida com um ser humano qualquer. Nesse sentido, Allan Turing propôs um teste que, em caso de sucesso, serviria para provar que uma máquina é inteligente. Esse teste é conhecido como Teste de Turing e analisa características como processamento de linguagem natural, representação de conhecimento, raciocínio automatizado, aprendizagem de máquina, além de visão computacional e a robótica como complemento.

Cabe apresentar cada um destes aspectos de forma sucinta, são eles:

- **Processamento de linguagem natural:** segundo Jurafsky e Martin (2019), o Processamento de Linguagem Natural (PLN) é uma área de pesquisa em Inteligência Artificial que emprega abordagens de aprendizado de máquina, estatística e linguística computacional para possibilitar que os computadores compreendam e produzam linguagem natural humana.
- **Representação de conhecimento:** para Brachman e Levesque (2004), a representação de conhecimento é um ramo da inteligência artificial que se dedica a como as informações podem ser simbolicamente representadas, armazenadas e recuperadas em sistemas computacionais, bem como essas informações podem ser manipuladas para gerar novos conhecimentos e conclusões sobre o mundo, por parte de agentes inteligentes.
- **Raciocínio automatizado:** no trabalho de Alan (1998), o raciocínio automatizado é definido como uma área da inteligência artificial que busca utilizar sistemas computacionais para realizar inferência lógica, aprendizado de máquina, planejamento e outras formas de raciocínio que possam ser automatizadas. O objetivo é permitir que as máquinas possam inferir novas informações a partir de dados já existentes e tomar decisões em situações onde há incertezas ou falta de informações.
- **Aprendizagem de máquina:** de acordo com Alpaydin (2020), aprendizagem de máquina é uma área da inteligência artificial que se concentra em criar algoritmos que possam aprender com dados e fazer previsões ou tomar decisões com base nesse aprendizado.
- **Visão computacional:** segundo Gonzalez e Woods (2008), a visão computacional engloba processos de aquisição, processamento e análise de imagens, com o objetivo de extrair informações e conhecimentos a partir delas.
- **Robótica:** Para LaValle (2006), a robótica é um campo da engenharia e da ciência da computação que se dedica ao desenvolvimento de robôs, máquinas capazes de executar tarefas de forma autônoma ou com supervisão humana. Ela engloba áreas como mecânica, eletrônica, inteligência artificial e controle de sistemas.

Este conjunto de habilidades, caso alcançadas pelo sistema artificial, poderia fornecer a classificação de inteligência por conta de sua operabilidade. O teste de Turing, então, não visa estabelecer se uma máquina é autoconsciente ou não e sim se ela opera conforme um comportamento inteligente esperado pelo padrão humano. Contudo, a Inteligência Artificial vai além da comparação com o ser humano e abarca aspectos voltados ao comportamento animal. Segundo Coppin (2004), a "inteligência artificial envolve o uso de métodos baseados no comportamento inteligente de humanos e outros animais para resolver problemas complexos".

Isso indica que a Inteligência Artificial agrega comportamento de animais, tais como abelhas e formigas. Os algoritmos de enxame de abelha e de fazenda de formigas são exemplos da modelagem de comportamento que utiliza o conceito de coletividade para resolução de

problemas de otimização. Nesses tipos de algoritmos, não é a inteligência em moldes humanos que se apresenta como foco, mas sim o comportamento instintivo que representa um grau de organização elevado.

No entanto, mesmo expandindo a concepção da Inteligência Artificial além da tentativa de reproduzir a inteligência do ser humano, trazendo também a tentativa de reproduzir o comportamento de animais, isto não é o suficiente para a percepção do que a Inteligência Artificial é capaz. Faz-se necessário apontar que existe uma subárea da Inteligência Artificial que diz respeito à uma linha evolucionária, ou seja, que se inspira inteiramente na seleção natural e na teoria da evolução de Darwin. Os algoritmos genéticos são exemplos dessa linha evolucionária e se enquadram como uma técnica de Inteligência Artificial porque eles são capazes de encontrar soluções para problemas de forma autônoma, sem intervenção humana direta. Eles imitam o processo evolutivo natural, criando e refinando soluções ao longo do tempo, até que uma solução satisfatória seja encontrada, tornando-os aplicáveis igualmente para resolução de problemas de otimização.

Diante deste amplo cenário, a Inteligência Artificial se mostra mais abrangente do que a reprodução de certos aspectos do ser humano. Entender esse espectro ampliado de possibilidades que essa tecnologia pode alcançar pode ajudar a compreender ou formular conceitos jurídicos sobre ela. Na Seção 3.2, são apresentados e discutidos os possíveis conceitos jurídicos da Inteligência Artificial, trazidos pelos Projetos de Lei 21/2020 e 2338/2023, os mais relevantes até o presente momento. Diante disto, é possível exercer uma visão crítica mais apurada sobre o texto legal e analisar seus pontos fortes e fracos.

Não obstante ao entendimento do que realmente é uma Inteligência Artificial, toda ela depende de três fatores: 1) do seu suporte, como as máquinas digitais; 2) da sua essência, do software e do ambiente digital; e 3) da autonomia tecnológica, que abstrai suas ações e reações da interferência humana. Esses elementos unidos fornecem um grau de inteligência operacional suficiente para a resolução de diversos tipos de problemas.

O grau de inteligência se traduz na potencialidade do que os sistemas inteligentes podem ou não realizar. Os modelos desenvolvidos, estudados e implementados até este trabalho não são capazes de evoluir por si, do mesmo modo que não detêm capacidade cognitiva para indicar autoconsciência. Neste sentido, há classificações que permitem agrupar e entender o estado atual da técnica de Inteligência Artificial, dentre elas está a dicotomia: 1) Inteligência Artificial fraca ou restrita; e 2) Inteligência Artificial Forte ou geral.

A abordagem da Inteligência Artificial Fraca parte do pressuposto de que os computadores podem auxiliar na compreensão da cognição humana, mas apenas por meio de simulações, sem que isto represente uma realização efetiva. Por outro lado, a Inteligência Artificial Forte defende que simular uma mente é equivalente a possui-la. Portanto, qualquer máquina que supere o Teste de Turing seria dotada de uma vida mental similar a do ser humano (LIMA FILHO, 2010). Entretanto, o atual estado da arte consegue simular, com certo grau de acurácia, a mente humana em atividades restritas e específicas, enveredando para uma

abordagem consistente com a Inteligência Artificial Fraca, pois esta caracteriza-se como uma operação simbólica, traduzindo-se em uma manipulação apenas sintática insuficiente para a essência genuinamente semântica existir.

Criar um programa de computador, mesmo que corretamente implementado, não é suficiente para que ele entenda o conteúdo semântico dos símbolos codificados através da sintaxe, pois não há intencionalidade nesta mera manipulação simbólica. O significado, neste contexto, é simulado por regras de reconhecimento de padrões e cálculos, escassos e exíguos, portanto, para constituir ou para produzir uma mente. Estados mentais não são apenas consequência de um comportamento, mas sim de uma conexão entre ele e uma estrutura causal, como o cérebro (LIMA FILHO, 2010). Desta forma, o cérebro é a origem fundamental da mente, o que entra em conflito com o Teste de Turing, que avalia as propriedades mentais de um sistema apenas por meio da observação de seu comportamento, ou seja, analisando as saídas geradas pelo sistema em resposta às entradas fornecidas pelo usuário humano.

A Inteligência Artificial é um software de computador e depende de sua programação e do nível de autonomia que lhe é fornecido por seus projetistas. Por isso, na Subseção 2.3.2, será apresentada a autonomia tecnológica e seus níveis de granulação. Isso se torna relevante, uma vez que quanto maior seu nível de autonomia, hipoteticamente seria maior seu grau de inteligência e, por consequência, maior será seu impacto no meio em que se está instalada.

2.3.2 Da autonomia tecnológica e da aprendizagem de máquina

A autonomia tecnológica está diretamente relacionada com o grau de afastamento do fator humano em determinadas atividades realizadas por sistemas artificiais, o que inclui os sistemas computacionais. Esse afastamento implica no autogerenciamento da própria tecnologia e, por consequência, das atividades por ela executadas. Para que essa autogerência se concretize, se faz necessária a implementação de uma camada de inteligência de software. Essa inteligência permite que atividades menores possam ser executadas de forma autônoma, atingindo o autogerenciamento como objetivo principal. Tais atividades, como atributos, são a autoconfiguração, a auto-otimização, a autocura e a autoproteção.

O atributo da autoconfiguração diz respeito ao sistema conseguir modificar, em tempo de execução, suas configurações de funcionamento. Essas mudanças permitem que o sistema altere seu comportamento de acordo com a percepção do ambiente externo do qual está diretamente ligado. Isso fornece resiliência ao sistema, sendo possível por conta do estabelecimento de uma rede de sensoriamento - de software, hardware ou ambos - que direciona as mudanças nos parâmetros de configuração do próprio sistema. Esse atributo é especialmente relevante quando se lida com sistemas que incorporam incerteza em seu cenário de execução, pois permite que se adapte as diversas condições que se apresentem.

O atributo da auto-otimização, por sua vez, tenta obter a configuração ótima para cada estado de ambiente que o sistema detecte, fazendo com que ele funcione de forma eficiente em todo seu ciclo de execução. A otimização, portanto, não consiste em uma mera mudança de

valores de configuração, mas sim em uma modificação que expresse o melhor funcionamento dentre das condições disponíveis em determinado momento.

Para o atributo da autocura, o sistema adquire a capacidade de conter danos ou mesmo de se regenerar ao ponto de continuar seu funcionamento com o mínimo esperado de qualidade, principalmente em situações de falha. Como reflexo, pode haver mudanças inesperadas na atuação do sistema e, naqueles que efetivamente tomam decisões e as executam, poderá existir a modificação significativa em seu fluxo principal de operação, afetando o resultado final consideravelmente.

Com o atributo de autoproteção, o sistema passa a ter a habilidade de se proteger contra investidas não autorizadas que podem resultar em acesso, modificação ou destruição indevida dos dados ou do sistema em si. A importância desse atributo está na garantia de segurança aguardada para a execução do sistema, que mitigaria ou eliminaria, em tese, distorções propositais vindas de fontes externas, isto é, a afetação deliberada do sistema por agentes externos mal-intencionados.

Todos os atributos mencionados, quando devidamente implementados, fornecem ao sistema o nível mais elevado de autonomia, o nível autonômico. Esse nível de autonomia foi proposto pela IBM (GANEK; CORBI, 2003), em conjunto com outros níveis menores. Esse conjunto é formato por cinco níveis e em cada um deles o ser humano tem alguma participação, mesmo que não diretamente. A Figura 7 apresenta os níveis de autonomia de acordo com a classificação da IBM.

Figura 7 – Níveis de autonomia.

Fonte: Adaptada de (GANEK; CORBI, 2003, p. 9).

Na Figura 7 é possível visualizar os níveis de autonomia de acordo com o frequência de interação humana e aprofundamento de especialização que são requeridas para cada um deles.

Esses níveis refletem a evolução da autonomia, iniciando no nível básico, onde a execução de todos os procedimentos de gerenciamento é de clara responsabilidade do ser humano até sistemas completamente autônomos.

Os níveis de autonomia do modelo apresentado podem ser resumidos da seguinte forma:

- **Nível Básico:** O sistema não possui automatização gerencial, restando para o ser humano a atividade de suporte em todos os aspectos. Esse deve possuir conhecimento especializado e aprofundado para tarefas como instalação, configuração, para adicionar ou remover componentes, estabelecer e manter a segurança da sistema, correção de falhas, dentre outras atividades essenciais.
- **Nível Gerenciado:** O sistema fornece suporte para sua gerência no formato centralizado. A principal característica desse nível está no monitoramento do contexto, que é apresentado por dados estatísticos. Além disso, há uma interface simplificada que permite que usuários não especialistas possam interagir e reconfigurar o sistema.
- **Nível Preditivo:** O sistema passa a monitorar seu próprio comportamento - além do contexto - e reconhece padrões, identificando tendências e apontando possíveis soluções. Essas serão apresentadas para o usuário administrador, que efetivamente será responsável por acatá-las ou rejeitá-las. Os sistemas que apresentam esse nível têm mecanismos de recomendação com base em correlações para identificação de mudanças que são necessárias em suas configurações.
- **Nível Adaptativo:** o sistema automatiza suas funções de gerenciamento, identifica anomalias e problemas através de reconhecimento de padrões e propõe soluções, aplicando diretamente algumas das soluções propostas sem a interferência do ser humano.
- **Nível Autonômico:** O sistema minimizará a necessidade de interação humana, executando suas atividades gerenciais de forma independente. Esses tipos de sistemas devem garantir os requisitos de qualidade, de confiabilidade e de segurança. O ser humano passa a ser responsável pelo estabelecimento de políticas de atuação do sistema, além de fiscalizá-lo. Usualmente, nesse nível, os sistemas possuem camadas de inteligências implementadas capazes de realizar a análise do próprio comportamento, escolhendo e decidindo quais ações serão tomadas para otimização de configurações, proteção e recuperação do sistema em caso de falhas.

Diante desta clara separação entre os níveis de autonomia aplicada a tecnologia digital, percebe-se uma relação diretamente proporcional entre os graus elevados de autonomia e a implementação de inteligência de sistema. Portanto, quanto mais inteligência é incorporada no sistema maior autonomia ele terá. Neste sentido, a Inteligência Artificial é elemento basilar para que a autonomia tecnológica avançada seja estabelecida. Assim sendo, a característica de extrapolação da representação de conhecimento que os sistemas inteligentes

têm, a aprendizagem de máquina, traz a resiliência necessária para sistemas alcançarem o nível autonômico, isto é, o maior nível de autonomia.

Por essa razão, a aprendizagem de máquina tem sido utilizada em sistemas que requerem adaptação e ações rápidas independentemente da vontade humana como matriz principal para seu funcionamento. A aprendizagem, no entanto, pode se apresentar como uma função simples ou como um dos desafios mais complexos e difíceis dentro da área de Inteligência Artificial, pois ela consiste no reconhecimento de padrões e na correlação entre dados como forma de aquisição de conhecimento.

Basicamente, quatro fatores devem ser observados quando se tem intenção de utilizar a aprendizagem de máquina em sistemas computacionais, são eles: 1) o que deverá ser melhorado; 2) o conhecimento prévio que o sistema já possui; 3) a representação utilizada para os dados e para os componentes; e 4) o *feedback* disponível para a aprendizagem. No primeiro, deve ser considerado o sistema, parte dele ou componente que será alvo do mecanismo de aprendizado para melhoramento de seu funcionamento. No segundo e no terceiro, o conhecimento já representado, ou seja, o conhecimento que o sistema já tem acesso. No quarto, os *feedbacks* podem ser classificados como aprendizagem supervisionada, aprendizagem não supervisionada, aprendizagem semi-supervisionada, aprendizagem por reforço e aprendizagem profunda.

Na aprendizagem supervisionada, de acordo com Alpaydin (2020), consiste em um tipo de aprendizado de máquina em que um modelo é treinado com exemplos rotulados para fazer previsões ou classificações em novos dados de entrada. O objetivo é que o modelo aprenda a mapear as entradas para as saídas esperadas a partir do conjunto de dados de treinamento. Essa abordagem permite o conhecimento das saídas do modelo, isto é, há a previsibilidade da resposta do sistema. Já para Goodfellow, Bengio e Courville (2016), a aprendizagem não supervisionada é um tipo de aprendizagem de máquina que não utiliza um conjunto de dados rotulados para treinar o modelo. Em vez disso, o algoritmo é alimentado com um conjunto de dados não rotulados e o objetivo é encontrar estruturas ou padrões significativos nesses dados sem conhecimento prévio sobre as classes ou categorias.

A aprendizagem semi-supervisionada, por sua vez, é uma abordagem de aprendizagem de máquina que utiliza tanto dados rotulados quanto não rotulados para realizar tarefas de classificação ou previsão. Essa abordagem visa melhorar a precisão do modelo ao incorporar mais informações disponíveis nos dados não rotulados. A ideia é que a incorporação de dados não rotulados, em conjunto com os dados rotulados, possa melhorar a capacidade de generalização do modelo (ZHU; GOLDBERG; ZHU, 2009). Para a aprendizagem por reforço se tem uma abordagem de aprendizado de máquina em que um agente aprende a tomar decisões através da interação com um ambiente, recebendo recompensas ou punições por suas ações. Essa abordagem é comumente usada em sistemas de inteligência artificial que envolvem tomada de decisão, como jogos, robótica e controle de processos industriais (SUTTON; BARTO, 2018).

A aprendizagem profunda é um subcampo da aprendizagem de máquina que utiliza redes neurais artificiais para modelar e solucionar problemas complexos de aprendizagem. Segundo

Goodfellow, Bengio e Courville (2016), a aprendizagem profunda envolve a construção de modelos compostos por várias camadas de unidades de processamento que aprendem a representação de dados de forma hierárquica e incremental. Esses modelos são capazes de aprender a partir de grandes conjuntos de dados para fazer previsões, identificar padrões e realizar tarefas complexas, como reconhecimento de fala, reconhecimento de imagens e processamento de linguagem natural.

Em todas as abordagens descritas, uma característica em comum: a complexidade em entender o processamento até a definição de uma saída como resultado. Isto se aplica também para as abordagens que se têm pleno conhecimento sobre as saídas do modelo. Isso porque mesmo que exista a previsibilidade em conhecer os "*outputs*" do sistema, não seria possível explicar com clareza como se deu o processamento até a chegada ao resultado entregue por ele.

Todo o cenário se torna mais complexo com a abordagem da aprendizagem profunda, pois os modelos utilizados são complexos e trabalham com massivo volume de dados, o *Big Data*[8]. Até a forma de se trabalhar com os dados mudam quando se tem *Big Data*. Entretanto, é possível extrair inteligência com grau de acurácia maior devido não só ao volume, mas também a velocidade de produção e coleta, e a variedade dos dados.

Nesse contexto de *Big Data*, é possível se obter um nível de autonomia máximo do sistema, pois este ficará menos dependente do ser humano como fonte de conhecimento primário para sua atuação. Portanto, na aprendizagem de máquina ou na aprendizagem profunda, a quantidade de dados para representação de conhecimento e aquisição de novos conhecimentos faz toda a diferença entre um sistema eficiente e ineficiente.

De qualquer forma, o ser humano ainda tem papel importante no funcionamento de sistemas inteligentes não importa o grau de autonomia. Esse papel diz respeito à fiscalização da atividade do sistema. Sua participação, no entanto, muda a depender dos níveis apresentados na Figura 7. Nos graus mais baixos, o ser humano é o núcleo do sistema, não apenas como um usuário passivo, mas como fonte das ações do sistema, ao passo que a autonomia se distancia das ações do ser humano, esse papel acaba for ser mitigado ao monitoramento apenas.

Entender o que é o ambiente digital e suas limitações, assim como compreender a relação entre este ambiente com o ambiente analógico, fornece uma visão aprofundada para o entendimento do que é um sistema de Inteligência Artificial. Isso porque esses tipos de sistemas ainda detêm como sua natureza de existência ser um software computacional alimentado por dados digitais, dos quais podem representar, também de forma limitada, qualquer coisa dentro do ambiente digital, inclusive pessoas.

[8]Big Data é um termo que se refere a grandes conjuntos de dados, estruturados ou não, que são complexos demais para serem processados por meio de ferramentas de processamento de dados tradicionais. De acordo com Chen et al. (2014), Big Data é caracterizado por ser volumoso, variável, veloz e de valor, exigindo o uso de técnicas computacionais avançadas para a sua análise e interpretação. O conceito de Big Data está relacionado com a ideia de que a capacidade de armazenamento e processamento de dados cresceu exponencialmente nos últimos anos, permitindo a coleta e armazenamento de uma grande quantidade de informações.

2.4 Dos direitos e garantias fundamentais e seus reflexos no ambiente digital

Os direitos fundamentais são essenciais para o ser humano e sua construção consiste em uma conquista histórica lentamente alcançada no decorrer do tempo. Por consequência disso, na literatura especializada, pode-se identificar dimensões das quais traduzem um incremento de direitos que se agregam como uma bola de neve que aumenta seu diâmetro ao percorrer certa distância, não existindo, neste sentido, o descarte de direitos anteriormente estabelecidos.

A primeira dimensão dos direitos fundamentais está relacionada às liberdades individuais focadas nos direitos civis e políticos, o que impõe um dever negativo para o Estado perante os indivíduos (DIÓGENES JÚNIOR, 2012). Na Constituição Federal, Brasil (1988), as liberdades se localizam no Art. 5º e listam uma série de direitos e garantias fundamentais, a exemplo dos direitos à privacidade, proteção de dados pessoas, à vida digna, à isonomia e equidade, à liberdade de expressão, dentre outros de igual importância. Todos os direitos estabelecidos na primeira dimensão são aplicáveis no ambiente digital, contudo alguns deles são mais perceptíveis do que outros em sua aplicação nesse ambiente.

A privacidade e a proteção de dados pessoais, assim como a liberdade de expressão parecem ser inerentes ao ambiente digital hiperconectado. Cada indivíduo, como usuário, deve ter sua privacidade preservada e apesar de ser possível rastrear suas passos e ações, é assegurada a não exposição em um ambiente onde todos poderiam ter acesso. De mesmo modo, cada titular de dados deve ter seus dados pessoais protegidos e utilizados somente através da observância aos direitos fundamentais e permissões normativas estabelecidas no ordenamento jurídico brasileiro. A liberdade de expressão, por sua vez, constitui um dos pilares basilares de construção da Internet como espaço de comunicação, sendo as redes sociais exemplos recentes de sua implementação no ambiente digital, como o *Twitter, Instagram, Facebook, Tik Toc* e muitas outras.

A segunda dimensão dos direitos fundamentais está ligada aos direitos sociais, econômicos e culturais - adequados aos direitos de igualdade - e, ao contrário dos direitos de primeira dimensão, impõe ao Estado um dever positivo a fim de garanti-los, isto é, ele deve agir para a efetividade destes direitos (PEREIRA, 2013). No Brasil, os direitos sociais estão no Art. 6º da Constituição Federal, trazendo o direito à educação, à saúde, ao trabalho, alimentação e transporte e outros. Sobre os direitos econômicos, o Art. 170 da constituição aponta a livre iniciativa como fundamento para se assegurar uma vida digna. Quanto aos direitos culturais, estabelecidos nos Arts. 215 e 216 da Constituição, eles são a base para o acesso à cultura nacional, visando a proteção da cultural.

Através do ambiente digital é possível obter formação educacional, o que ficou em evidência com o surgimento da pandemia da Covid-19 em 2020. A modalidade de Educação à distância já se fazia presente na sociedade, porém existiu uma situação de migração forçada para ambiente digital devido ao estado de confinamento compulsório para combate à pandemia. Com isto, a modalidade de educação telepresencial se fez presente de maneira intensa na realidade

dos brasileiros. Na saúde, o atendimento médico pode ser realizado através do digital com a utilização da Internet, possibilitando o acesso quando a situação assim o permite. Do mesmo modo que ocorrera com a educação, as relações de trabalho sofreram impacto importante com a pandemia, possibilitando o uso da modalidade remota como alternativa ao presencial. Quanto ao transporte e alimentação, as tecnologias digitais trouxeram ferramentas que permitem realizar pedidos de entrega de comida, assim como para chamar um transporte para deslocamento.

A livre iniciativa fez emergir novos modelos de negócio fundados inteiramente em ambiente digital ou realizáveis por intermédio dele. As plataformas de *Streaming*, como o Spotify, Disney+, Netflix e um vasto espectro de serviços semelhantes são exemplos destes novos modelos aplicados à indústria do entretenimento. Serviços bancários e de crédito também podem ser listados neste rol, a exemplo do Nubank, Banco Inter, etc.

A terceira dimensão dos direitos fundamentais diz respeito aos direitos que detêm estreitamento com a solidariedade e a fraternidade. Nestes direitos, não há uma delimitação precisa em seus limites, por consequência, o molde da titularidade não tem uma configuração precisão (RODRIGUES, 2013). Dentre os direitos relacionados à essa dimensão estão o direito ao meio ambiente ecologicamente equilibrado, direito ao progresso sustentado, direito à autodeterminação dos povos, direito do consumidor e o direito à comunicação.

Para o consumidor, hipossuficiente nas relações de consumo e exposto em sistemas complexos e distanciados do fornecedor, a vulnerabilidade é ressaltada naqueles que não possuem afinidade com a digital e podem ser lesados diante de plataformas digitais irregulares. Já para o direito à comunicação, a Internet implementa esse direito no ambiente digital e através dela as mais variadas formas de comunicação podem ser efetivadas, desde mensagens instantâneas e vídeos, até às cartas eletrônicas, ou seja, os e-mails.

Apesar da existência não pacifica na literatura especializada quanto aos direitos representados na quarta dimensão dos direitos fundamentais, estes compreendem os direitos à informação, à democracia, ao pluralismo político (SANTOS, 2010), manipulação genética e estão relacionados à globalização (SILVA, 2009). Neste plano, o digital impacta no desenvolvimento de tecnologias que aceleram o progresso dos estudos na área de genética e biologia, assim como aperfeiçoam o uso da informação como mola propulsora da democracia. A informação decorre de dados e evoluem para conhecimento, sendo através deste a possibilidade do alcance da inteligência. Atualmente, uma quantidade vasta e significativa de dados estão armazenados em ambiente digital. Além disto, softwares automatizam a utilização de dados e a construção da informação, o que impacta a pesquisa em diversas áreas, inclusive a biológica e a genética.

Há quem defenda uma quinta e uma sexta dimensão dos direitos fundamentais. Na quinta dimensão, a cibernética e a informática são ressaltadas, em conjunto com o direito à paz (OLIVEIRA, 2016). Na sexta dimensão estaria o direito à água potável como fonte primordial da vida e fundamental para sobrevivência do ser humano (SANTOS; PINTO; ADAME, 2016). A própria descrição de cibernética e informática remete ao ambiente digital, enquanto à paz e

a água potável é possível fazer relação com o digital, tanto por unir Estados e povos para o diálogo, aproximando culturas, como também para monitoramento do acesso à água potável e outras possíveis vertentes de aplicação.

Em todas as dimensões, mesmo que não reste pacificada quanto aos direito correspondentes, o ambiente digital se faz presente e influencia na aplicação e efetividade dos direitos fundamentais no mundo moderno. O digital emergiu, em vários aspectos, com vantagens e desvantagens enquanto interação humana dentro desse ambiente artificial. O software materializa os instrumentos de concretização dos direitos fundamentais no digital e a utilização de sistemas inteligentes elevou o impacto e as consequências, isto é, o risco aos direitos fundamentais, positivamente e negativamente.

Assim, a constitucionalização do meio ambiente digital pode influenciar a proteção dos direitos fundamentais de duas maneiras fundamentais. Primeiramente, ela pode ampliar as oportunidades para as pessoas exercerem seus direitos, como a liberdade de expressão. No entanto, ao mesmo tempo, pode também aumentar os riscos para esses direitos, incluindo questões de privacidade e segurança cibernética. Portanto, as respostas constitucionais devem considerar esses efeitos positivos e negativos para garantir uma proteção adequada dos direitos fundamentais no meio ambiente digital (CELESTE; SANTARÉM, 2021).

Desta forma, o termo, segundo o trabalho de Celeste e Santarém (2021), "constitucionalismo digital" reconhece o papel que a tecnologia digital tem desempenhado recentemente como o principal catalisador da mudança no ambiente constitucional. Em particular, permite-nos distinguir o ramo específico ou declinação do constitucionalismo moderno que está traduzindo e adaptando valores e princípios constitucionais existentes às peculiaridades da sociedade digital contemporânea. Os autores apontam que o termo "constitucionalismo digital" é frequentemente usado para descrever a resposta constitucional aos desafios da tecnologia digital. No entanto, eles observam que não existe uma definição única desse conceito na literatura acadêmica.

Segundo os autores citados no parágrafo anterior, propõem uma classificação das medidas normativas desenvolvidas em resposta aos desafios apresentados pela tecnologia digital em três categorias principais. Primeiramente, tem-se as normas destinadas a reconhecer a capacidade ampliada de exercer direitos fundamentais já existentes. Em segundo lugar, encontram-se as normas voltadas para a proteção de direitos fundamentais que enfrentam ameaças devido à tecnologia digital. Por fim, temos as normas criadas para regular fenômenos novos originados da tecnologia digital, que não se encaixam nas categorias anteriores. Além disso, ressalta-se que as respostas constitucionais às mudanças trazidas pela tecnologia digital não se limitam ao direito constitucional tradicional. Elas abrangem normas de direito ordinário, direito constitucional e documentos de direitos específicos para a internet. Essas respostas transcendem os limites nacionais e podem ser observadas em níveis regionais e internacionais, bem como em mecanismos de resolução de disputas de organizações transnacionais, como a *Internet Corporation for Assigned Names and Numbers* (ICANN).

Não se pode esquecer que as normas constitucionais possuem a característica de imperatividade, que é um atributo presente em todas as normas jurídicas. Caso tais normas não sejam cumpridas, os mecanismos de coação e cumprimento forçado devem ser acionados (BARROSO, 2005). É imperativo, porquanto, estabelecer o entendimento que as normas constitucionais, em especial os direitos e garantias fundamentais, devem ser respeitados no ambiente digital assim como no ambiente analógico.

Essa força normativa da Constituição Federal, da qual se manifesta com a concretização dos direitos e garantias fundamentais no ambiente digital por meio de softwares, produzindo efeitos reais para os indivíduos, também influencia outras normas jurídicas que estão sob sua validação. A partir da elevação da Constituição a um papel central no sistema jurídico, houve a concretização da chamada constitucionalização do direito infraconstitucional. Esse processo é caracterizado pela supremacia material da Constituição, cuja normatividade é expressa pelos seus princípios (SANTOS, 2021). Por essa razão, todo o ordenamento infraconstitucional, para a real concretização nesse ambiente através do software, também necessitará estar em consonância com a constituição, restando como requisito para sua efetivação, primeiramente a verificação de conformidade com a constituição e somente depois desta validação existir a tradução para um artefato de software.

Além disso, a concretização da Constituição Jurídica está diretamente relacionada à sua habilidade de antecipar o futuro tendo em vista as circunstâncias do presente. Dessa forma, a força e a eficácia da Constituição se baseiam na sua conexão com as forças espontâneas, tendências vitais da sociedade e o tempo em que é aplicada (LIMA; LANÇA, 2013). Isso significa que a Constituição precisa ser capaz de se adaptar às mudanças e às necessidades da sociedade ao longo do tempo, de modo a continuar a ser relevante e eficaz. Portanto, para que a Constituição seja efetiva, ela deve estar em constante diálogo com a realidade social, econômica e política do país. Como tais realidades estão em expansão para o ambiente digital, se torna natural que a constituição exerça sua força normativa igualmente neste ambiente.

Por isso, espera-se que a Constituição Brasileira de 1988 vá além de regulamentar as questões do Estado, inclusive no ambiente digital, possuindo uma amplitude que se estende à própria sociedade. Sua aplicação é exigida em situações específicas, gerando efeitos até mesmo em relações entre particulares, nas quais o Estado não possui envolvimento direto (BERNARDI; PIEROBON, 2014), o que incluiria o microssistema de proteção de dados pessoais.

Neste cenário, a Constituição passou a exercer influência na interpretação da legislação infraconstitucional, o que alguns chamam de filtragem constitucional. Isso afetou diversos ramos do direito, incluindo o direito privado, que era normatizado principalmente pelo Código Civil. Como resultado, a Constituição passou a ser vista como um guia para a interpretação de todas as normas do ordenamento jurídico brasileiro (SANTOS, 2021). Não somente isso, a elevação de normas de natureza civil para a constituição é um movimento que aponta essa constitucionalização do direito privado. Um exemplo desse fenômeno é a proteção de dados pessoais, disciplinada especificamente por lei infraconstitucional, tendo natureza civil, e por

força da Emenda Constitucional 115 foi alçada à Constituição. Esse movimento será melhor explorado no Capítulo 3 , Seção 3.1.

No Capítulo 3 será discutido como o Direito enfrenta o tema de tratamento de dados pessoais realizados por Inteligência Artificial, a começar pela proteção dos dados até a regulamentação dessa tecnologia no Brasil, perpassando pela responsabilização civil por danos causados por ela.

3 DO USO DA INTELIGÊNCIA ARTIFICIAL E DA PROTEÇÃO DE DADOS PESSOAIS

Neste capítulo, o principal objetivo reside no entendimento da relação entre o direito fundamental à proteção de dados pessoais com o uso de sistemas de software que implementam técnicas de Inteligência Artificial. De mesma forma, procurou-se entender brevemente os impactos que o uso dessa tecnologia tem nos direito e garantias fundamentais e as tentativas de regulamentação do uso de Inteligência Artificial no Brasil. Buscou-se também trazer reflexões sobre a responsabilização daqueles que projetam e utilizam sistemas inteligentes.

Entender a relação entre o direito fundamental à proteção de dados pessoais com a utilização de sistemas com Inteligência Artificial é importante porque esse tipo de tecnologias é empregada constantemente e de forma abrangente na sociedade atual, impactando-a de maneira sem precedentes. Por essa razão, se fez necessária a compreensão das características dessa tecnologia trazidas no Capítulo 2. Isso é fundamental para entender que a tecnologia, apesar de seu potencial, é limitada por conta do ambiente em que se está, do qual também é limitado.

Por conta desta limitação e das características inerentes da tecnologia em questão, discute-se brevemente os impactos do uso da Inteligência Artificial trazendo seus benefícios e malefícios para o indivíduo e para a sociedade, ajudando a concretizar direitos por um lado e por outros podendo violá-los. Após, a discussão sobre as tentativas de regulamentação da Inteligência Artificial no Brasil através de projetos de lei sobre o tema, com evidência aos projetos que podem se tornar o Marco Legal da Inteligência Artificial no Brasil e suas nuances.

Por fim, em decorrência dos impactos negativos significativos, traz-se também a discussão sobre responsabilidade civil sobre o uso da Inteligência Artificial, no tocante a identificar os atores de desenvolvimento e uso em ambiente real, levando-se em consideração os níveis de autonomia tecnológica implementada no sistema, níveis esses apresentados no Capítulo 2 na Subseção 2.3.2.

A metodologia utilizada coaduna com a revisão de literatura narrativa, com as palavras-chave específicas sobre os temas citados conforme o protocolo estabelecido no Apêndice B.

3.1 Do direito fundamental à proteção de dados e os requisitos para o tratamento de dados por Inteligência Artificial

Os dados refletem e representam objetos, pessoas e relações dentro do ambiente digital. Através dos dados é possível identificar padrões, criar perfis, traçar tendências, obter inteligência para agir sobre certas circunstâncias, registrar e comprovar transações e até mesmo perpetuar a própria história. Os dados sempre compuseram a estrutura de um software, sendo processados e apresentados como resposta ao ser humano.

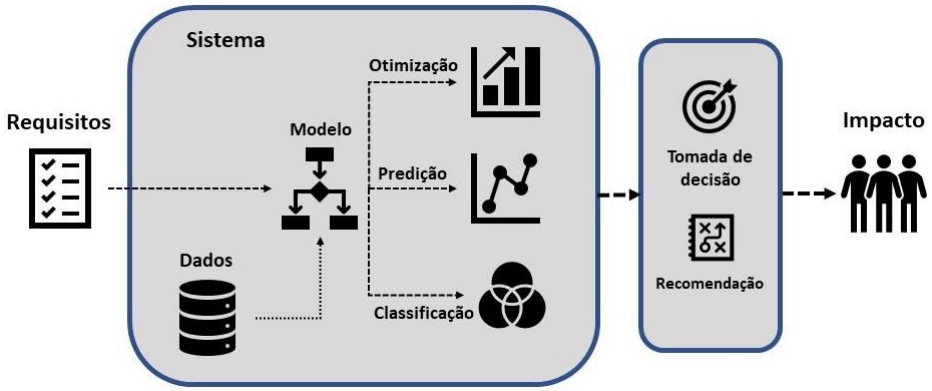
Além disso, os dados agora servem para nutrir sistemas de Inteligência Artificial de forma que estes possam se aperfeiçoar e evoluir, tornando-se resilientes. Esses sistemas

inteligentes são construídos com alto nível de especificação, cujo intuito consiste em usufruir dos dados digitais com grau elevado de acurácia e eficiência em áreas de aplicação específicas. Inevitavelmente, dados pessoais são incluídos no tratamento digitalizado e automatizado desses sistemas.

Isso porque os dados são símbolos sem significado relevante quando analisados isoladamente. Porém, ao serem agrupados com outros dados que possuem alguma relação, é possível enriquecer seu valor semântico e formar uma informação. Em outras palavras, os dados são elementos constituintes da informação, e seu processamento resulta em informação. Quando um contexto mais amplo é adicionado à informação, é possível adquirir conhecimento, o que é fundamental para a inteligência e a tomada de decisão (SIRIHAL; LOURENÇO, 2002). No mesmo sentido, Barbieri (2019) aponta que o contexto de inteligência se traduz em tomada de decisão, perfilamento, otimização e outros modos de interferência na sociedade e para os indivíduos que a compõe.

Os modelos de aprendizagem de máquina utilizam dados para treinamento, teste e para pleno funcionamento, em ambientes de produção. No entanto, os dados precisam ser preparados antes de sua introdução nos modelos de inteligência artificial, pois, caso contrário, podem levar a resultados discriminatórios ou com violação de direitos, que estão longe de representar o cenário esperado. A Figura 8 expõe uma arquitetura simplificada dos modelos com aprendizagem de máquina.

Figura 8 – Arquitetura Simplificada de sistema de Inteligência Artificial com aprendizagem de máquina.



Fonte: O autor.

A Figura 8 ilustra a interação entre os requisitos de funcionamento do sistema, definidos pelos idealizadores do sistema, o modelo concebido pelos projetistas e engenheiros de inteligência artificial, os dados utilizados para treinamento e teste, as saídas geradas (atividades exemplificadas na Figura 2 e com os requisitos apontados na Figura 3) - que neste exemplo incluem classificação, predição e otimização - e, por fim, a interferência no ambiente analógico, impactando tanto o indivíduo quanto a sociedade, através de ações como a tomada de decisão ou no direcionamento de comportamentos com sistemas de recomendação.

Esse tipo de tratamento de dados pessoais possui probabilidade de danos consideráveis

para aqueles que são submetidos a ele. Esses danos podem aumentar significativamente quando o tratamento é realizado por sistemas digitais. No entanto, os danos causados por sistemas de Inteligência Artificiais são imensuráveis, devido ao alcance volumétrico do tratamento realizado.

Os impactos individuais, coletivos e sociais do uso de Inteligência Artificial são perceptíveis e expressivos. Esses impactos ocorrem, possivelmente, pela utilização descontrolada deste tipo de tecnologia. Particularmente, nesses sistemas, há uma relação diretamente proporcional entre poder computacional e consumo de dados. Portanto, quanto maior o poder computacional maior será o consumo de dados. Entretanto, o que permitiu o avanço da tecnologia foi a quantidade e variedade dos dados disponíveis atualmente (LUGER, 2013).

A disponibilidade mencionada representa, em relação aos dados pessoais, uma exposição de seus titulares - aqueles aos quais pertencem tais dados. Isso torna o titular vulnerável ao tratamento de dados por si só, pois os resultados do tratamento afetam diretamente a concretização de seus direitos, inclusive os fundamentais.

A movimentação para a proteção destes dados pelo legislador se tornou mais evidente pelo infraconstitucional, com a elaboração da Lei nº 12.965 de 2014, conhecida como Marco Civil da Internet e dispõe sobre princípios, garantias e deveres para a utilização da Internet no Brasil (BRASIL, 2014). Neste dispositivo legal, há o apontamento para a proteção de dados pessoais no artigo 3º, inciso III, do qual traz os princípios inerentes ao uso da Internet. Todavia, a expressão "na forma da lei", que acompanha tal inciso, deixava o regramento para o uso desses dados para lei posterior.

Em 2018, a Lei nº 13.709 de 2018 foi introduzida no ordenamento jurídico brasileiro e estabeleceu princípios, finalidades, deveres e responsabilidades para aqueles que tratam dados pessoais, além de trazer direitos e garantias para os titulares de dados. A Lei Geral de Proteção de Dados tem como principal objetivo a regulamentação da utilização de dados pessoais por pessoas jurídicas e naturais que utilizam estes mesmos dados com fins lucrativos, protegendo, deste modo, a parte hipossuficiente da relação de tratamento de abusos e violações de direitos.

A importância da proteção de dados obteve tamanha relevância ao ponto de alçar ao *status* constitucional através da PEC 17/2019, que resultou na Emenda Constitucional de nº 115 de 2022, adicionando ao rol dos direitos e garantias fundamentais o direito à proteção de dados pessoais, contido no artigo 5º, LXXIX da Constituição brasileira, o qual diz que "é assegurado, nos termos da lei, o direito à proteção dos dados pessoais, inclusive nos meios digitais" (BRASIL, 1988). A proteção de dados, a partir de então, adquire, de certa forma, um desenho autônomo na esfera constitucional. Neste ponto, a proteção de dados agora se irradia do núcleo do sistema jurídico normativo e se propaga por todas as esferas do direito.

O direito à proteção de dados pessoais se consagra como uma liberdade individual, trazendo todas as características inerentes dela, a começar pela aplicabilidade imediata, por força do § 1ª do artigo 5º da Constituição Federal de 1988, pois já se insere com lei infraconstitucional

que regula a matéria. A Emenda Constitucional referida, além de reconhecer a proteção de dados como direito fundamental, também estabelece regulamentações em nível constitucional. Para isso, foi adicionado ao artigo 21 da Constituição Federal o inciso XXVI, que atribui à União a responsabilidade de "organizar e fiscalizar a proteção e o tratamento de dados pessoais, de acordo com a lei" (BRASIL, 1988). Por sua vez, o artigo 22 da Constituição foi alterado, incluindo o inciso XXX, que estabelece que somente a União pode legislar sobre a proteção e o tratamento de dados pessoais. A Emenda ainda estabelece a exclusividade da União em termos de competência para legislar sobre o tema.

A valorização da proteção de dados pessoais no cenário dos direitos fundamentais ressalta sua relevância ao lado de outros direitos, como a privacidade, o sigilo das comunicações, a dignidade da pessoa humana e outros, contribuindo para a construção de uma sociedade mais livre, justa e respeitosa. A proteção de dados pessoais e a privacidade são interdependentes e complementares. Assim, a proteção de dados pessoais é uma parte importante da privacidade, pois o uso inadequado de informações pessoais pode violar a privacidade das pessoas. Todavia, a privacidade é fundamental para a proteção de dados pessoais, uma vez que a coleta, processamento, armazenamento e compartilhamento de informações pessoais sem autorização pode expor a intimidade e a vida privada dos indivíduos (SOMBRA, 2019).

Ambos os direitos fundamentais visam proteger os indivíduos contra o uso indevido de suas informações pessoais, permitindo-lhes manter o controle sobre seus dados. A proteção de dados pessoais permite que as pessoas decidam quando, como e para que fins suas informações pessoais são tratadas. Já a privacidade permite que os indivíduos mantenham sua vida privada longe do público ou de terceiros não autorizados. A proteção de dados pessoais e a privacidade trabalham em conjunto para garantir que as pessoas tenham controle sobre suas informações pessoais e sua vida privada, permitindo-lhes desfrutar de uma maior autonomia e dignidade (SARLET, 2020). Ambos os direitos são fundamentais para o exercício de outros direitos, como a liberdade de expressão, a liberdade de associação e o direito à igualdade, por exemplo.

A proteção de dados pessoais e o sigilo das comunicações são fundamentais para garantir que os indivíduos possam exercer seus direitos sem a interferência indevida de outras pessoas ou entidades e são cruciais para a proteção da democracia e do Estado de Direito. No entanto, em algumas situações, esses dois direitos podem entrar em conflito, como quando as autoridades públicas precisam acessar dados pessoais ou interceptar comunicações para fins de segurança nacional ou investigação criminal. Nesses casos, é importante buscar um equilíbrio adequado entre esses dois direitos para garantir a proteção da privacidade e segurança dos indivíduos ao mesmo tempo em que se permite a proteção da segurança nacional e a prevenção do crime.

No tocante à relação entre a dignidade da pessoa humana e a proteção de dados pessoais, observa-se que ambos são essenciais para garantir que cada indivíduo possa exercer seus direitos e viver com dignidade e liberdade, pois a dignidade da pessoa humana é um princípio fundamental que reconhece que cada pessoa possui um valor intrínseco, independentemente de sua origem, raça, sexo, religião ou outra característica, enquanto a proteção de dados

pessoais, por sua vez, permite que cada indivíduo possa controlar tais informações pessoais. Isso é essencial para a proteção da privacidade e da liberdade dos indivíduos, que são valores fundamentais para a dignidade da pessoa humana.

Destarte, qualquer tratamento que envolva dados pessoais deve observar primeiramente a constituição a fim de não violar quaisquer direitos e garantias fundamentais, então a Lei Geral de Proteção de Dados, por trazer requisitos para o devido tratamento, e os demais regramentos específicos que regulem a utilização de dados pessoais devem observá-la rigorosamente. Portanto, a automatização destes tratamentos, por consequência, deve igualmente seguir os preceitos constitucionais e legais estabelecidos. Isso significa que a automatização realizada por software e sistemas de computadores precisam estar adequados e em conformidade desde a concepção destes artefatos e permanecer até sua depreciação, isto é, quando o software não será mais utilizado. Dessa forma, esse tipo de tratamento aumenta significativamente o nível de risco para o titular dos dados, tanto em termos positivos quanto negativos. E quando se emprega Inteligência Artificial, esse risco é elevado a um patamar ainda mais relevante.

O grau de autonomia tecnológica concedido a um sistema de Inteligência Artificial tem um impacto direto nos efeitos que esse sistema pode ter sobre as pessoas, autonomia esta apresentada na Subseção 2.3.2. Por essa razão, deve-se considerar cuidadosamente o nível de autonomia concedido aos sistemas inteligentes e suas possíveis consequências para os indivíduos afetados. Quanto maior o nível de autonomia tecnológica maior deverá ser seu monitoramento para que o funcionamento da tecnologia não produza tratamento em desconformidade com os requisitos legais e constitucionais (AFFONSO et al., 2019).

Com a implementação de grau elevado de autonomia fica evidente a necessidade da observação dos princípios trazidos pela Lei Geral de Proteção de Dados, em seu artigo 6ª. Em especial, os princípios da finalidade, da adequação e necessidade, assim como o da transparência, não discriminação e responsabilidade, prestação de contas e prevenção. A finalidade na utilização do sistema de Inteligência Artificial precisará de clareza e especificação, devendo ser informada de maneira explícita ao titular que haverá um tratamento automatizado por sistemas deste tipo. De mesmo modo, é preciso estabelecer e esclarecer a necessidade de tratamento dos dados que serão coletados, pois o motor de raciocínio e inteligência do sistema será movido por esses dados. Já o princípio da adequação diz respeito ao encaixe entre a finalidade e necessidade do uso da ferramenta inteligente e caso não exista uma convergência entre eles o tratamento se tornará inadequado.

Garantir a não discriminação na utilização dos sistemas automatizados consiste em um dos principais objetivos na proteção do indivíduo perante tais tecnologias. Isso porque o resultado discriminatório causa prejuízos ao indivíduo, o que enseja responsabilidade e dever de reparação. A consequência da responsabilização é possível através da transparência, que terá discussão mais aprofundada no Seção 4.1. Neste aspecto, não se pode esquecer da qualidade dos dados porque este princípio garante que os dados estejam sanitizados, isto é, preparados para que seu tratamento não direcione em risco de enviesamento com resultados discriminatórios ou que

violem direitos mitigado.

O princípio da prevenção aponta a necessidade em adotar medidas preventivas com o fim de evitar que ocorra qualquer dano ou violação dos dados pessoais dos titulares. Em outras palavras, o princípio da prevenção exige que as entidades que tratam dados pessoais implementem medidas técnicas e organizacionais adequadas para garantir a proteção dos dados desde o momento da coleta até o armazenamento e descarte. Isso coaduna com os demais princípios, quando o objetivo é diminuir a probabilidade de ocorrer prejuízos ao titular e a sociedade. Pode-se, neste ponto, fazer um comparativo com o princípio da prevenção vinda do Direito Ambiental, que também tem o intuição de exigir que sejam tomadas medidas preventivas antes que ocorram danos irreversíveis ao meio ambiente (CIELO et al., 2012). O cerne em ambos está em agir proativamente para evitar possíveis danos, uma vez que nos dois existem incertezas científicas sobre os efeitos de determinada atividade, uma quanto ao meio ambiente e outro quanto aos titulares de dados.

Mesmo quando se observam os princípios para o tratamento de dados pessoais, é importante destacar que esse tratamento não pode ser realizado de forma livre e desimpedida. A Lei Geral de Proteção de Dados estabelece permissões para que esse tratamento seja realizado, incluindo sistemas com Inteligência Artificial que processam dados pessoais. Dentre essas permissões, o consentimento, contido no artigo 7º, inciso I da Lei Geral de Proteção de Dados, parece ser uma base residual, utilizada de forma subsidiária quando o tratamento de dados não tiver nenhuma outra permissão que se adéque melhor para a situação.

Permissões legais que se evidenciam com o uso de sistema de Inteligência Artificial constam no artigo 7º, inciso V e X, do mesmo dispositivo legal citado no parágrafo anterior, dos quais falam sobre contrato e proteção ao crédito respectivamente. Em ambos, sistemas inteligentes são utilizados para análise e concessão de créditos, seguros, a formação de contrato para assegurar o cumprimento do acordo estabelecido entre as partes envolvidas, para a emissão de faturas, entrega de produtos ou prestação de serviços de suporte pós-venda, por exemplo. Nestes cenários, o processamento e a resposta dele são fornecidos quase que de forma instantânea e, por isso, os efeitos desse processamento podem ser sentidos também de maneira instantânea.

Para o Poder Público, as permissões legais mais relevantes são a de cumprimento de obrigação legal ou regulatória, presente no artigo 7º, inciso II, a do tratamento pela administração pública, artigo 7º, inciso III, o legítimo interesse, artigo 7º, inciso IX e o consentimento como base residual para o tratamento, no artigo 7º, inciso I, todos trazidos na Lei Geral de Proteção de Dados. Essas permissões legais possibilitam a introdução de sistemas automatizados dentro das atividades do Poder Público nos três poderes. No Capítulo 5 discute-se sobre a utilização de sistemas com Inteligência Artificial dentro do processo judicial para as mais diversificadas atividades dentro dos procedimentos que compões os atos processuais, incluindo-se o auxílio à tomada de decisão. Outro ponto de interesse para o Estado é a execução de políticas públicas, onde os sistemas inteligentes são empregados em situações também variadas. Na pandemia do

COVID-19, a tecnologia foi implementada para identificação de padrões de aglomerações de pessoas para que houve ações de dispersão.

A mais frágil das permissões legais para tratamento automatizado é o consentimento, pois sua simples remoção acarreta na impossibilidade de uso dos dados do titular que retirou sus permissão. Em casos em que se têm milhões de dados para alimentação e extrapolação de conhecimento talvez o impacto não se configure significativo, porém se os titulares massivamente optem por retirar essa permissão, o motor de raciocínio das Inteligências Artificiais será comprometido em seu funcionamento eficaz.

Em todo o caso, os titulares de dados precisam ter ciência desse tipo de tratamento. Essa ciência deve ser baseada em informações relevantes, discussão focada no Capítulo 4. A razão da preocupação em informar ao titular sobre o uso de seus dados nestes artefatos complexos de software reside na possibilidade do próprio titular exercer seus de direitos, em especial os elencados no artigo 17 e Art 20 da Lei Geral de Proteção de Dados.

Dentre os direitos apontados no parágrafo anterior, os que se destacam por conta da relação com tratamentos automatizados são o direito ao acesso, o direito à correção, o direito à exclusão, o direito de portabilidade, os já mencionados direito de informação e direito de revogação de consentimento, além dos direitos à revisão de decisões automatizadas e o direito à explicação, foco desta pesquisa. O direito de acesso e o direito de informação se complementam diretamente no sentido de trazer a ciência do tratamento, permitindo que o titular saiba exatamente que dados são coletados, produzidos a partir dos dados coletados e o compartilhamento destes com sistemas de Inteligência Artificial.

O direitos à correção e o direito à exclusão dos dados fazem referência à qualidade e sanitização dos dados do titular e que formam a base de conhecimento que o sistema utilizará para sua execução. Isso se faz necessários, uma vez que dados incompletos, inexatos e desatualizados causarão impacto significativo no sistema. De mesmo modo, a exclusão de dados de um titular que não deseja fazer parte dessa base de conhecimento deve ser garantida, observando-se a boa-fé e a ética que repousam na construção da base de dados. Neste último, há exceção para a manutenção desses dados, mesmo em casos que o titular requeira sua exclusão, caso sejam necessários para o cumprimento de obrigações legais ou regulatórias. Quanto ao direito à revogação de consentimento, nos casos em que o titular retirar o consentimento fornecido previamente, os dados deverão ser excluídos da mesma maneira. Essa operação de exclusão deverá ser executada porque o mero armazenamento é considerado um tratamento e como não existirá mais a permissão para tratar o dado, ele não deve existir mais na base de conhecimento.

O titular detém o direito à portabilidade de seus dados, sendo este um direito dependente da interoperabilidade entre sistemas. O intuito aqui está em transladar dados de um sistema para outro. Essa portabilidade não poderá de nenhuma forma atrapalhar os direitos e liberdades de terceiros. Já o direito à revisão de decisões automatizadas e o direito à explicação dizem respeito ao poder que o titular de dados tem de não se submeter de forma irrestrita ao tratamento

automatizado e desumanizado. Contudo, no Brasil, não se obriga a revisão de decisões ou perfilamento automatizado por seres humanos, o que pode provocar um cascateamento de sistemas inteligentes processando dados e revisando como foi realizado o tratamento do sistema anterior. O direito à explicação será abordado no Capítulo 4.

Todo o tratamento de dados pessoais executado por Inteligência Artificial deve, portanto, observar os requisitos legais e constitucionais, como mencionado nesta seção. Não se deve temer o uso da tecnologia além do receio necessário para uma atuação responsável desses sistemas junto aos indivíduos e a sociedade. A análise dos impactos sobre os direitos e garantias fundamentais será discutido a seguir.

3.1.1 Breve análise dos impactos aos direitos e garantias fundamentais advindos da utilização de Inteligência Artificial

Os sistemas automatizas por Inteligência Artificial funcionam a partir de uma base de conhecimento, seja representado a partir de especialista ou construídos a partir de dados apenas. A abordagem que utiliza a representação de especialista tem como objetivo traduzir o conhecimento do profissional especialista no intuito de resolver determinado problema dento do domínio de conhecimento dele.

No entanto, caso o domínio do sistema precise de alteração, a exemplo de um sistema de Inteligência Artificial médico para um sistema de Inteligência Artificial jurídico, a representação realizada no primeiro caso (por um especialista médico) é ineficaz e ineficiente para servir para um sistema de Inteligência Artificial jurídico, necessitando a representação, então, de um profissional da área jurídica. Esta mudança tem grau de investimento significativo do ponto de vista financeiro, de tempo e de programação. Além do mais, o conhecimento representado também é limitado ao conhecimento do profissional especialista, sendo uma desvantagem em termos de extrapolação de conhecimento.

Com a abordagem a partir dos dados, o próprio sistema aprenderá com eles e com o contexto em que está inserido. Não há necessidade da pessoa do especialista e agora dos dados pode-se alcançar a inteligência. Para a mudança de domínio de funcionamento, basta somente alterar o conjunto dos dados para um conjunto mais adequado (ADAMSSEN, 2020).

Os dados são fundamentais para os atuais modelos de aprendizagem de máquina utilizados e os dados pessoais possuem relevante destaque. Os dados pessoais representam seu titular no ambiente digital, uma vez que é possível identificá-lo, os tendo como ponto de partida. Esses dados alimentam os sistemas de Inteligência Artificial, em seus *inputs*, restando os *outputs* para a finalidade do sistema. Todo o processamento feito pela Inteligência Artificial significa um tratamento de dados automatizado. Implica dizer que esse tipo de tratamento atinge uma quantidade massiva de pessoas e esta exposição pode ter efeitos positivos e negativos em termos de privacidade e proteção de dados pessoais, liberdade de expressão, acesso à educação, na seara da saúde, nas atividades laborais e isto se espalha por todos os direitos fundamentais.

Sistemas de Inteligência Artificial podem garantir a privacidade e proteção de dados,

quando aplicados à segurança de dados e informação, garantindo que os titulares de dados estejam protegidos. Contudo, o tratamento desregrado sem a devida observância da legislação específica viola o direito à proteção de dados pessoais, à privacidade e, por consequência, outros direitos fundamentais. Para a liberdade de expressão, os sistemas de Inteligência Artificial podem suprimir tal direito, controlando as comunicações dos usuários, limitando o que acessam e o que publicam em ambiente digital compartilhado, como a Internet.

Na educação, a Inteligência Artificial pode auxiliar na aprendizagem dos discentes recomendando conteúdo de acordo com as necessidades das disciplinas e suas preferências. Além disto, sistemas de Inteligência Artificial podem decidir direcionar verbas para colégios, universidades e mesmo definir quem receberá bolsa de pesquisa ou não. Na seara trabalhista, Inteligência Artificial pode ser aplicada em processos de seleção para vagas de emprego, decidindo quem está apto ou não para seguir no processo seletivo. Diante de dados pessoas, como endereço, idade, gênero e cor da pela, o sistema poderá eliminar candidatos que residam longe do local de trabalho, que tenham idade superior aquela desejada, eliminar mulheres e diversas outras consequências imprevisíveis.

Na saúde, os sistemas de Inteligência Artificial auxiliam no diagnóstico médico e auxiliam, igualmente, em cirurgias remotas, limitando possíveis excessos cometidos pelo profissional médico por equívocos. Em relação aos transportes e pedidos de alimentos, a Inteligência Artificial influencia em limitação de disponibilidade de lugares e motoristas, fazem a ponte com sistemas financeiros e outros serviços de suporte.

Já para acesso à crédito, esses sistemas, assim como acontece com a seleção para contratação laboral, podem com base nos dados brutos negar a concessão de crédito financeiro e para o setor de seguro ocorre o mesmo. São esses mesmos tipo de sistemas que escolhem quem receberá promoção de um determinado produto e que não receberá, mesmo que não exista diferença nítida entre consumidores neste sentido.

Um dos principais problemas encontrados no uso da Inteligência Artificial está na compreensão do funcionamento dos atuais modelos de aprendizagem de máquina, cuja lógica de execução é incompreensível. Isto quer dizer que não há uma explicação facilitada do como determinada ação efetuada por esses sistemas foi realizada. Na ausência de entendimento do processo de tratamento automatizado, não há garantia de efetividade do quaisquer direitos.

Contudo, nos casos em que se há no tratamento automatizado por Inteligência Artificial dados pessoais, a Lei Geral de Proteção de Dados traz garantias de que exista uma explicação deste tratamento, artigo 20, §1º - "O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial" (BRASIL, 2018), e que esta seja significativa para o titular de dados. Desta forma, o direito à explicação implementaria o princípio da transparência, igualmente definido na Lei Geral de Proteção de Dados, em seu Art 6º, 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" (BRASIL, 2018).

O projeto de lei PL 20/2020 que visa regulamentar o desenvolvimento e o uso de Inteligência Artificial no Brasil, em seu artigo 5º, V, traz como princípio a transparência, do qual diz: transparência: "direito das pessoas de serem informadas de maneira clara, acessível e precisa sobre a utilização das soluções de inteligência artificial, salvo disposição legal em sentido contrário e observados os segredos comercial e industrial..." (BRASIL, 2020, p. 4), ainda no inciso V, alínea c, tem-se "sobre critérios gerais que orientam o funcionamento do sistema de inteligência artificial assegurados os segredos comercial e industrial, quando houver potencial de risco relevante para os direitos fundamentais" (BRASIL, 2020).

Há, de fato, uma preocupação com o desenvolvimento e a utilização de sistemas de Inteligência Artificial, pois seu resultado pode ser incerto e ocasionar danos imensuráveis. A legislação tentará direcionar o rumo desta tecnologia sem engessá-la. Contudo, cabe fiscalizar seu uso constantemente e garantir que possíveis erros possam ser corrigidos e danos reparados. Na Seção 3.2 será discutida a tentativa e regulamentação da tecnologia de Inteligência Artificial no Brasil.

3.2 A regulamentação da inteligência artificial no Brasil e sua relação com a proteção de dados

A regulamentação da Inteligência Artificial tem sido objeto de discussões em todo o mundo, inclusive no Brasil. Em 2019, foi criado um grupo de trabalho para elaborar propostas de regulamentação da Inteligência Artificial no país, que resultaram em um documento preliminar divulgado em agosto de 2020. Em abril de 2021, foi lançada uma consulta pública sobre a regulamentação da Inteligência Artificial no Brasil, com o objetivo de receber contribuições da sociedade para o aprimoramento do texto.

Devido a importância dos impactos da Inteligência Artificial, evidencia-se que a regulamentação da Inteligência Artificial no Brasil é fundamental para garantir a proteção dos indivíduos e da sociedade diante dos potenciais riscos e impactos negativos dessa tecnologia. Por isso, é essencial que haja transparência e ética na utilização desses sistemas automatizados, a fim de garantir que não haja discriminação ou violação de direitos fundamentais.

Além disso, a coleta e o processamento de dados pessoais por meio da Inteligência Artificial podem comprometer a privacidade e a segurança das informações dos indivíduos. Por isso, é importante que a regulamentação da Inteligência Artificial no Brasil inclua medidas de proteção de dados pessoais, como a obrigatoriedade de consentimento informado, a transparência na utilização desses dados e a garantia de direitos aos titulares dessas informações (PEIXOTO; COUTINHO, 2020).

Assim, a regulamentação da Inteligência Artificial no Brasil deve estabelecer critérios claros e transparentes para o uso dessa tecnologia, garantindo que tais princípios éticos e os direitos humanos sejam respeitados. Entre as questões que devem ser abordadas na

regulamentação da Inteligência Artificial estão o processo de tomada de decisão, a prevenção contra a discriminação negativa, a proteção da privacidade e da segurança dos dados pessoais, a responsabilização das empresas e dos desenvolvedores por danos causados pela Inteligência Artificial, entre outros diversos aspectos relevantes.

Essa regulamentação, no Brasil, deve ser um processo participativo e democrático, que envolva a sociedade, as instituições privadas, os desenvolvedores e os especialistas no tema. É importante que as normas estabelecidas sejam adequadas ao contexto brasileiro, levando em consideração as especificidades sociais, econômicas e culturais do país.

Há várias propostas legislativas para a regulamentação do tema em tramitação no congresso federal até a produção desta pesquisa. Os Projetos de Lei de números 21 de 2020 (BRASIL, 2020), 5.051 de 2019 (BRASIL, 2019a), 5691 de 2019 (BRASIL, 2019b), 872 de 2021 (BRASIL, 2021) e o 2338 de 2023 (BRASIL, 2023) são exemplos das tentativas de regulamentação da Inteligência Artificial, sendo o primeiro e o último os mais relevantes. O Projeto de Lei 21 de 2020 tem por objetivo estabelecer fundamentos, princípios e diretrizes para o desenvolvimento e para a aplicação de Inteligência Artificial no Brasil, já o 2338 visa estabelecer normas gerais para o desenvolvimento, implementação e uso para sistemas com Inteligência Artificial em território nacional, priorizando a proteção de direitos fundamentais e a garantia de sistemas seguros e confiáveis.

O Projeto de Lei 5051 de 2019 também estabelece princípios para o uso da tecnologia, além de fornecer diretrizes para a atuação da União, Estados, Distrito Federal e Municípios para o desenvolvimento de sistemas inteligentes. O Projeto de Lei 5691 de 2019 vem com a proposta de estabelecer a Política Nacional de Inteligência Artificial, cujo objetivo é a estimulação e a formação de uma ecossistema que seja favorável ao desenvolvimento desta tecnologia. Igualmente, o Projeto de Lei 872 de 2021 dispõe sobre Inteligência Artificial no Brasil, trazendo objetivos da utilização da tecnologia e diretrizes para o Estado desenvolvê-la e utilizá-la.

Diante da não especificação da expressão “decisão automatizada” dentro da Lei Geral de Proteção de Dados, houve a tentativa de descrever com detalhes o que seria uma decisão automatizada no contexto de proteção de dados pessoais com o Projeto de Lei 4496 de 2019 (BRASIL, 2019c). Com o projeto, seria adicionado o inciso XX ao artigo 5º da referida lei para adicionar o conceito de decisão automatizada como uma etapa de seleção, avaliação, pontuação, classificação, cálculo de riscos ou probabilidade, aprovação ou rejeição, ou outras atividades similares, em que os dados pessoais são processados por meio de regras, instruções, algoritmos, análises estatísticas, inteligência artificial, aprendizado de máquina ou outras técnicas computacionais.

Essa definição trazida pelo Projeto de Lei 4496 de 2019 poderia ajudar a definir o contexto em que uma decisão automatizada se relacionaria com a Inteligência Artificial, pois não se faz necessária, a priori, que um sistema automatizado se utilize de técnicas mais avançadas como essa para a tomada de decisão porque uma simples lógica tradicional de Se-Então-Senão poderia resolver o problema da decisão, mas de forma dura e sem flexibilidade.

Isto posto, tem-se então projetos de lei que visam definir e trazer os caminhos para o desenvolvimento e aplicação da tecnologia, não somente no setor privado como também no público, no território brasileiro. Dentre os projetos citados, o mais relevantes em termos de discussão são os Projetos de Lei 21 de 2020 e 2338 de 2023, dos quais serão objetos para análise nesta seção. Neles o objetivo em comum consiste em regular o uso da Inteligência Artificial no Brasil, buscando garantir que as tecnologias sejam desenvolvidas e utilizadas de maneira ética e responsável, de modo a proteger os direitos e interesses dos indivíduos e da sociedade como um todo.

O projeto 21 de 2020 abrange tanto o setor público como o privado, estabelecendo diretrizes para a utilização da Inteligência Artificial em diversas áreas como na saúde, na educação, no judiciário, na segurança, entre outras. O projeto ainda apresenta um conceito jurídico de Inteligência Artificial baseado no prisma da Ciência da Computação, no artigo 2º, definindo como sistema de Inteligência Artificial aquele que opera por meio de processamento computacional e é capaz de aprender a reconhecer e interpretar o ambiente externo, além de interagir com ele, com base em um conjunto de objetivos definidos por humanos. Esse sistema poderá realizar previsões, oferecer recomendações, classificações ou tomar decisões utilizando dados e informações e podem incluir técnicas como: I) sistemas de *Machine Learning*, que englobam técnicas de aprendizado supervisionado, não supervisionado e por reforço; II) sistemas que tem como base a lógica ou em representação de conhecimento; e III) outras técnicas utilizadas incluem o uso de análises estatísticas, inferência bayesiana, métodos de pesquisa e de otimização.

O parágrafo único do artigo 2º, no entanto, limita o alcance da futura lei indicando que ela não é válida para os processos de automação que são dirigidos apenas por parâmetros preestabelecidos de programação, os quais não contêm a habilidade do sistema em aprender a perceber e interpretar o ambiente externo e interagir com ele com base nas ações e informações recebidas. Em outras palavras, decisões automatizadas com técnicas tradicionais não estão incluídas na regulamentação do projeto.

No artigo 3º do Projeto de Lei 21 de 2020 são fixados os objetivos para a aplicação da Inteligência Artificial, tendo como principal o desenvolvimento científico e tecnológico. Pode-se observar, diante do texto presente neste artigo, que o intuito é a promoção e o fomento no desenvolvimento e uso da tecnologia para um desenvolvimento econômico sustentável e para bem-estar social, como disposto no inciso I desse artigo. A inserção do Brasil como um ator relevante no cenário internacional referente ao tema resta claro com os incisos I e III. Com isso, indica-se que a regulamentação trará, na realidade, um incentivo à pesquisa e desenvolvimento e não terá intenção de engessar ou atrasar o desenvolvimento tecnológico, nesse sentido.

Já no artigo 4º do projeto, existe uma ratificação da intenção de incentivo e promoção deste tipo de tecnologia. Neste artigo, são mostrados os fundamentos para o desenvolvimento e para a aplicação de Inteligência Artificial no país. Dentre eles estão a livre iniciativa, a livre concorrência, o incentivo à autoregulamentação e a liberdade dos modelos de negócios. Nesse

sentido, a Inteligência Artificial pode ser vista como uma ferramenta para impulsionar a livre iniciativa, uma vez que a automação de processos pode aumentar a eficiência, reduzir custos e melhorar a qualidade dos produtos e serviços, pois a livre iniciativa se refere à liberdade de atuação dos indivíduos e empresas na busca por seus objetivos econômicos, sem a interferência do Estado em sua atividade econômica.

Por outro lado, a livre concorrência está diretamente ligada à manutenção de um mercado justo, com igualdade de oportunidades para todas as empresas que desejam competir. A Inteligência Artificial pode ser utilizada como uma forma de nivelar o campo de jogo, pois empresas de todos os tamanhos podem ter acesso a tecnologias avançadas de automação e processamento de dados. Além disso, a Inteligência Artificial pode ser usada para garantir uma concorrência leal, por exemplo, por meio de sistemas de detecção de fraudes e abusos de mercado.

Entretanto, é importante lembrar que a utilização da Inteligência Artificial também pode gerar desigualdades, como a concentração de poder e recursos nas mãos das empresas que possuem tecnologias mais avançadas. Por isso, a regulamentação adequada da Inteligência Artificial é fundamental para garantir um mercado justo e equilibrado para todas as empresas e para a sociedade como um todo.

A autoregulamentação da Inteligência Artificial pode trazer benefícios para o desenvolvimento e uso da tecnologia no Brasil, pois permite que as empresas e organizações sejam responsáveis por suas próprias práticas de Inteligência Artificial, ao invés de dependerem de regulamentações governamentais rígidas e inflexíveis. Através da autoregulamentação, as empresas podem criar padrões éticos e práticas responsáveis de Inteligência Artificial que se adéquem ao seu contexto e setor de atuação. Além disso, a autoregulamentação pode ser mais rápida e eficiente na adaptação às mudanças e avanços tecnológicos do que as regulamentações governamentais.

Todavia, a autoregulamentação também pode apresentar desvantagens. Como não haveria uma supervisão externa, as empresas podem ser tentadas a colocar seus interesses financeiros acima dos interesses da sociedade e de seus usuários. Além disso, a autoregulamentação pode levar à fragmentação de padrões éticos e práticas responsáveis de Inteligência Artificial, dificultando a adoção de um conjunto unificado de regulamentos. Por fim, no inciso VII, ainda no artigo 4º, existe tentativa de prender a autoregulamentação aos princípios trazidos pelo Projeto de Lei e por boas práticas globais que pode mitigar as desvantagens apresentadas.

Outro ponto sensível está na proteção à livre manifestação de pensamento e a livre expressão da atividade intelectual, artística, científica e de comunicação, contido no inciso IV do Art 4º, do projeto de lei. A Inteligência Artificial tem um potencial significativo para proteger e promover tais liberdades. No entanto, é importante que a tecnologia seja usada de maneira responsável e equilibrada, a fim de evitar efeitos prejudiciais para esses liberdades. Isso porque a Inteligência Artificial pode ajudar a proteger e promover esses direitos, permitindo que mais

pessoas possam criar e compartilhar suas ideias e trabalhos com o mundo. Por exemplo, sistemas de Inteligência Artificial podem ser usados para ajudar a identificar violações de direitos autorais em plataformas online, permitindo que os criadores de conteúdo sejam recompensados pelo uso de seu trabalho.

Por outro prisma, os sistemas de Inteligência Artificial também podem ser usados para limitar a livre manifestação de pensamento e a livre expressão. Algoritmos que contém censura podem ser usados como justificativa para restringir o acesso a conteúdo considerado ofensivo ou prejudicial, que na realidade são legítimos e importantes para o debate público. Além disso, a Inteligência Artificial pode ser usada para criar conteúdos falsos ou enganosos, o que pode minar a confiança nas informações disponíveis online e prejudicar a capacidade das pessoas de tomar decisões informadas.

A não discriminação, a pluralidade, o respeito às diversidades regionais, a inclusão e o respeito aos direitos e garantias fundamentais do cidadão são fundamentos trazidos no inciso V ainda referente ao artigo 4º d projeto. Pautar o desenvolvimento e uso de sistemas de Inteligência Artificial no fundamento da não discriminação garante que a tecnologia não seja usada para perpetuar preconceitos, estereótipos ou discriminação baseada em raça, gênero, orientação sexual, religião, idade ou outras características. A pluralidade deve, da mesma forma, ser levada em conta na criação de soluções desse tipo, garantindo a diversidade de perspectivas e pontos de vista que reflitam a sociedade atual. Em razão disso, esses fundamentos devem ser levados em consideração em fase de projeto, desde de seu início, quando se levantam os requisitos de negócio, como apresentado no Capítulo 2, na Seção 2.2.

O projeto estabelece um conjunto de princípios éticos que devem guiar o desenvolvimento e uso da Inteligência Artificial no país, tais como: a finalidade benéfica, cujo foco está nos benefícios para a humanidade; a centralidade do ser humano, o que coloca o respeito à dignidade da pessoa humana e seus os direitos fundamentais sempre como prioridade quando os sistemas tratarem de questões que se relacionam com o ser humano; a transparência, fornecendo ao ser humano informações sobre o sistema (a transparência será discutida com maior profundidade no Capítulo 4); a responsabilidade, fazendo com que os atores que atuam na cadeia de existência da Inteligência Artificial respondam por danos causados; a segurança, que traz a obrigação da utilização de todas as medidas, dentro da viabilidade da ocasião, técnicas e organizacionais para mitigar riscos advindos da operação dos sistemas inteligentes em todo o ciclo de sua existência; a não discriminação, como abordado nesta seção, como forma de eliminar resultados discriminatório no processo de tratamento de dados; e entre outros fundamentos pertinentes.

Além disto, traz consigo uma série de diretrizes para a aplicação da Inteligência Artificial, como a gestão baseada em risco, a participação social e interdisciplinar, a análise de impacto regulatório e a responsabilidade. Essa preocupação em mitigar os riscos no uso dos sistemas inteligentes está no fato de que as tecnologias modernas incorporarem no tratamento automatizado que realizam o aspecto da incerteza de resultados devido aos cenários complexos

e dinâmicos que estão inseridos. Por isso, deve-se levar em consideração os riscos concretos no desenvolvimento e uso desses tipos de sistemas. Isso significa que a análise dos riscos deve ser feita de forma detalhada e criteriosa, levando em conta as possíveis consequências negativas que o sistema pode gerar, o que inclui, segundo o projeto, os benefícios socioeconômicos com o uso e a comparação com os riscos de sistemas similares que não agregam técnicas de Inteligência Artificial.

Já a previsão da participação social e interdisciplinas encontra-se em convergência com os interesses do ser humano em não se expor além do necessário para essa tecnologia. Portanto, deve-se garantir que aqueles que serão inevitavelmente afetados por ela possa participar de maneira ativa e efetiva na criação de normas que visem regulamentar esta matéria. A importância dessa premissa é a garantia que a regulamentação da Inteligência Artificial seja baseada em evidências concretas e na opinião de todos os envolvidos, especialistas e a sociedade. Isso aumenta a probabilidade de que as normas adotadas sejam eficazes e adequadas às necessidades e desafios apresentados pela tecnologia. A premissa também enfatiza a importância da participação democrática no processo de regulamentação. Isso ajuda que as normas sejam aceitas como legítimas perante a sociedade, contribuindo para a construção de um ambiente seguro e confiável para o desenvolvimento e uso da Inteligência Artificial.

O projeto de lei em questão inclui a exigência de uma análise de impacto regulatório para regulamentar a Inteligência Artificial. Esse processo envolve uma avaliação de risco legal e regulatório para determinar as diretrizes necessárias para essa regulamentação. O artigo 6º, inciso V, menciona a Lei 13.874 de 2019 e o Decreto 10.411 de 2020, que fornecem as orientações necessárias para realizar essa análise de impacto. Então, antes de criar os atos normativos, será necessário realizar um procedimento que envolve a avaliação prévia do problema regulatório. O objetivo desse procedimento é analisar os possíveis efeitos do ato normativo, utilizando informações e dados, para determinar se o impacto é razoável e subsidiar a tomada de decisão.

Outro ponto importante do projeto está na previsão do regime de responsabilidade civil para os agentes que atuam na cadeia de existência (desenvolvimento e operação), isto é, para aqueles que desenvolvem ou utilizam o sistema de Inteligência Artificial, de modo a garantir que esses atores reparem possíveis danos causados pelo sistema. Nesse ponto, o Art 6º, inciso VI atribui a responsabilidade subjetiva para esses atores, considerando-se a efetiva participação de cada um deles dentro da cadeia de existência da Inteligência Artificial. O projeto define a culpa como elemento nuclear para responsabilização e sua dosimetria consta na extensão e gravidade dos danos que deveria evitar ou que terá que remediar, na devida adequação com às normas aplicáveis à sua atividade e o emprego de esforços compatíveis com padrões internacionais e boas práticas.

Após essa breve análise do projeto 21 de 2020 segue, a partir daqui, também uma breve análise de projeto 2338 de 2023, que aborda a Inteligência Artificial sob uma perspectiva centrada no ser humano e na proteção dos valores democráticos e direitos humanos. Isso

se justifica devido aos consideráveis riscos e impactos significativos associados aos sistemas dinâmicos de Inteligência Artificial. Este projeto incorpora uma série de princípios para garantir essa proteção, alguns dos quais se assemelham aos apresentados no Projeto de Lei 21/2020, como os princípios da não discriminação e transparência. Esse projeto é um espelhamento da legislação europeia sobre Inteligência Artificial, o *EU IA Act*, em um movimento semelhante do que ocorreu com a Lei Geral de Proteção de Dados e o Regulamento Geral de Proteção de Dados Europeu.

O projeto também traz o conceito de sistema de Inteligência Artificial, levando-se em conta os graus diferentes de autonomia, o que diferencia dos demais conceitos trazidos pelos demais projetos apontados aqui. O projeto também fala expressamente sobre representação de conhecimento além de aprendizagem de máquina, o que alarga as possibilidade de adequação do conceito aos sistemas inteligentes. O conceito direciona que a atividade desses tipos de sistemas é alimentada inicialmente por dados advindos de máquinas ou humanos, porém em muitos casos as duas interfaces produzem ou coletam tais dados para entrada no sistema. Há, contudo, objetivos elencados que servem como exemplos das saídas produzidas pelo sistema, são elas: as predições, as recomendações ou decisões que podem, de alguma forma, influenciar o ambiente digital e o real.

O projeto trouxe, neste ponto em específico referente à finalidade, ao final do artigo 4º, inciso I, as saídas objetivadas dos sistemas, abstraindo as atividades de perfilamento, classificação, inferência e regressão e a otimização como técnicas implementadas. Isso possibilita a perenidade do conceito sem sua obsolescência. Como a tecnologia computacional e, por consequência, as técnicas de Inteligência Artificial, tende a evoluir rapidamente, é desejável que a norma jurídica a acompanhe não restando tão precisa ao ponto de perder essa conexão.

O projeto em questão define novos atores envolvidos no ecossistema de Inteligência Artificial, do qual vai desde o desenvolvimento até a operação dos sistemas, são eles: o fornecedor de sistema de Inteligência Artificial, aquele responsável pelo desenvolvimento do sistema tanto diretamente ou por encomenda, e o operador de sistema de Inteligência Artificial, aquele que empregue ou utilize o sistema em seu nome ou em seu benefício, ambos como agentes de Inteligência Artificial; e a autoridade competente, que deverá ser um órgão ou entidade da administração pública federal que fiscalizará o cumprimento da lei. Essa estrutura se assemelha ao proposto pela Lei Geral de Proteção de Dados, que contém os agentes de tratamento e, de igual forma, uma autoridade nacional de proteção de dados.

O destaque desse projeto está na categorização dos graus de riscos associados aos sistemas com Inteligência Artificial. Os graus definidos no projeto são os de Risco Excessivo e o Risco Alto expressamente, ignorando os graus menos impactantes como o médio e o baixo, incluindo-os de forma implícita, normalmente atribuídos em sistemas de gradação de riscos. Inicialmente, a responsabilidade pela categorização caberá ao fornecedor da tecnologia, através de uma análise de risco voltada à Inteligência Artificial, da qual contará com a avaliação contínua de impacto algorítmica, prevista na Seção III, do projeto.

Em relação ao risco dito como excessivo, o projeto é explicitamente proibitivo, o que resulta no impedimento da implementação de sistemas com este nível de risco. No artigo 14 do projeto, os incisos I, II e III apontam, de forma exemplificativa, hipóteses de sistemas inteligentes presumidamente de risco excessivo. No inciso I, não é permitido o uso de técnicas subliminares com o objetivo de induzir pessoas a adotarem comportamentos prejudiciais à saúde ou segurança. Também veda, em seu inciso II, a exploração de vulnerabilidades específicas, como idade ou deficiência, para induzir tais comportamentos. Além disso, no inciso III, impede o poder público de avaliar, classificar ou pontuar pessoas com base em seu comportamento ou atributos de personalidade de forma ilegítima ou desproporcional para o acesso a bens, serviços e políticas públicas.

No contexto das atividades de segurança pública, contido no artigo 15, somente seria autorizado o emprego de sistemas de identificação biométrica à distância, como o reconhecimento facial, de maneira contínua em locais públicos, mediante previsão em lei federal específica e aprovação judicial vinculada à investigação penal individualizada. A utilização desses sistemas será permitida apenas para a persecução de crimes com pena máxima superior a dois anos, na busca de vítimas de crimes ou pessoas desaparecidas, ou em casos de crime flagrante. Conforme indicado pelo parágrafo único deste artigo, a legislação correspondente deve definir medidas que estejam em equilíbrio com o interesse público, assegurando o devido processo legal, supervisão judicial e aderência aos princípios desta Lei. Isso inclui a prevenção da discriminação e a revisão da inferência algorítmica antes que qualquer medida seja tomada em relação à pessoa identificada.

No contexto do projeto, os sistemas de alto risco podem ser implementados para finalidades específicas, de acordo com as categorias delineadas no artigo 17. Esses sistemas estão sujeitos a medidas de governança adicionais, além das aplicáveis a sistemas de menor risco, discutidos na Seção I do Capítulo IV do projeto. Tais medidas envolvem a necessidade de documentação técnica prévia à disponibilização ou uso do sistema, a manutenção dessa documentação ao longo da operação e a adoção de medidas para garantir transparência e mitigação de vieses prejudiciais aos direitos e garantias.

A documentação técnica deve ser elaborada abarcando a avaliação de impacto algorítmico é essencial, identificando e mitigando possíveis riscos e efeitos negativos nos direitos e liberdades fundamentais. Sistemas de alto risco impõem medidas adicionais de governança. Por fim, sistemas governamentais de inteligência artificial devem seguir normas específicas que garantam transparência, imparcialidade e proteção dos direitos dos cidadãos.

A avaliação de impacto algorítmico consiste em um processo sistemático e documentado que visa identificar, avaliar e mitigar possíveis riscos e efeitos negativos em direitos e liberdades fundamentais decorrentes da utilização de sistemas de inteligência artificial e está definido na Seção III ainda no Capítulo IV, seguindo do artigo 22 até 26 do projeto 2338.

Os critérios e a metodologia para realizar a avaliação de impacto algorítmico envolvem diversas etapas. Primeiramente, é necessário definir o escopo da avaliação, identificar os

envolvidos e os riscos associados ao sistema de inteligência artificial. Em seguida, ocorre a identificação e análise dos riscos, incluindo aqueles conhecidos e previsíveis à época do desenvolvimento, bem como os que razoavelmente podem ser esperados. Posteriormente, medidas são definidas e implementadas para mitigar os riscos identificados, incluindo ações para evitar ou minimizar vieses e discriminações, garantindo equidade e imparcialidade. Por fim, são estabelecidos indicadores e mecanismos de monitoramento para avaliar a eficácia das medidas adotadas e a possível ocorrência de novos riscos.

A avaliação de impacto algorítmico deve ser conduzida por profissionais ou equipes com conhecimentos técnicos, científicos e jurídicos adequados para a elaboração do relatório, assegurando independência funcional. A autoridade competente também pode estabelecer critérios adicionais, incluindo a participação dos diversos segmentos sociais afetados, de acordo com o risco e o porte econômico da organização em questão.

A autoridade competente deve atualizar regularmente a lista de sistemas de inteligência artificial de alto risco, com base em critérios como implementação generalizada, impactos negativos nos direitos e liberdades, potencial prejudicial e discriminatório, efeitos sobre grupos vulneráveis e a irreversibilidade de resultados danosos. Esses sistemas englobam diversas finalidades, como segurança em infraestruturas críticas, educação, recrutamento, análise de crédito, saúde, acesso a serviços públicos e privados, estratégias de investimento, tomada de decisões judiciais e administrativas, entre outras, como indicado nos incisos I a XIV do artigo 17. Durante todo o ciclo de vida desses sistemas, medidas de governança devem ser aplicadas, incluindo transparência para permitir compreensão e verificação, bem como mitigação de vieses para garantir equidade.

Em relação à responsabilidade civil, o texto do projeto estabelece a responsabilidade civil objetiva para os fornecedores ou operadores de sistemas de Inteligência Artificial classificados como de alto risco ou risco excessivo. Isso significa que eles são responsáveis pelos danos causados independentemente de culpa, sendo sua responsabilidade determinada com base na medida de sua contribuição para o dano, como diz o artigo 27 §1º. No caso dos sistemas de inteligência artificial que não se enquadram como de alto risco, a culpa do agente causador do dano será presumida, o que implicaria na inversão do ônus da prova em favor da vítima, como estabelecido no §2º do mesmo artigo.

O uso de sistemas de Inteligência Artificial pelo Poder Público, especialmente aqueles classificados como de alto risco, requer a implementação de medidas específicas de governança, conforme já mencionado anteriormente. Além disso, o texto estabelece diretrizes adicionais para orientar o uso responsável desses sistemas.

Isso inclui a realização de consultas e audiências públicas anteriores à utilização planejada dos sistemas, com divulgação de informações detalhadas sobre os dados a serem empregados, a lógica geral de funcionamento e resultados de testes prévios. Também é fundamental estabelecer protocolos de acesso e uso dos sistemas, incluindo a necessidade de registrar quem os utilizou, em que contexto e com qual finalidade. É importante que

os dados utilizados sejam provenientes de fontes seguras, precisos, relevantes, atualizados e representativos das populações afetadas, sendo testados para evitar vieses discriminatórios.

Além disso, deve ser garantido o direito do cidadão a uma explicação e revisão humanas das decisões tomadas por sistemas de Inteligência Artificial, com facilidade de acesso perante o poder público. Por fim, são necessárias medidas que assegurem a transparência, a imparcialidade e a proteção dos direitos fundamentais dos cidadãos, incluindo a realização de avaliações de impacto algorítmico e a definição de normas específicas de governança para sistemas de Inteligência Artificial governamentais. Estas diretrizes têm como objetivo primordial promover a transparência, a equidade e a proteção dos direitos essenciais dos cidadãos no contexto do uso de sistemas de Inteligência Artificial pelo Poder Público.

Diferente de outros projetos de lei sobre o tema, no 2338 de 2023, é definia uma autoridade fiscalizadora. A autoridade competente desempenhará várias funções cruciais para o fomento da pesquisa, desenvolvimento e inovação em inteligência artificial, além de estabelecer diretrizes essenciais para a aplicação efetiva da Lei. Adicionalmente, essa entidade deverá manter um fórum contínuo de comunicação com os órgãos e entidades responsáveis pela regulamentação de setores específicos da atividade econômica e governamental. Isso visa aprimorar a coordenação das suas competências regulatórias, de fiscalização e de aplicação de sanções.

É relevante destacar que os regulamentos e normas elaborados pela autoridade competente passarão por um processo de consulta e audiência públicas, além de análises de impacto regulatório. Essas medidas visam promover a transparência, a participação pública e a avaliação abrangente dos efeitos das regulamentações propostas, garantindo que sejam eficazes e equitativas em seu escopo.

No que diz respeito ao *sandbox* regulatório, contido no artigo 38 do projeto, o texto estabelece que a autoridade competente tem a prerrogativa de conceder autorização para o funcionamento de um ambiente experimental regulatório voltado para a inovação em inteligência artificial, conhecido como *sandbox* regulatório, às entidades que cumpram os requisitos estipulados pela Lei e sua regulamentação. As solicitações para a obtenção de autorização para esses *sandboxes* regulatórios devem ser apresentadas ao órgão competente por meio de projetos que evidenciem, entre outras coisas, a inovação na utilização da tecnologia ou em alternativas ao uso de tecnologias já existentes, melhorias que visem a aumentar a eficiência, reduzir custos, elevar a segurança, diminuir riscos e proporcionar benefícios para a sociedade e os consumidores.

O propósito do *sandbox* regulatório reside em proporcionar às entidades a oportunidade de testar novas soluções em um ambiente controlado, com flexibilidade regulatória, para avaliar a viabilidade e eficácia dessas soluções antes de sua implementação em larga escala. Isso visa a estimular a inovação, garantir a segurança e a eficácia das tecnologias de inteligência artificial e promover benefícios tangíveis tanto para a sociedade em geral quanto para os consumidores.

No contexto das sanções administrativas, contido no Capítulo VII, Seção II, artigo 36,

o texto estabelece diversas medidas que a autoridade competente pode impor aos agentes de Inteligência Artificial em caso de infrações às normas desta Lei. Essas sanções, contidas nos incisos de I até VI, abrangem advertências, multas proporcionais ao faturamento das empresas, publicização das infrações, restrições à participação em regimes de *sandboxes* regulatório, suspensão das operações de sistemas de Inteligência Artificial e a proibição do tratamento de determinadas bases de dados. A aplicação das sanções ocorrerá após a condução de um procedimento administrativo que garanta o direito à ampla defesa e levará em conta as circunstâncias específicas de cada caso, podendo ser aplicadas de forma isolada ou cumulativa, conforme determinado pelos critérios estabelecidos na Lei.

Em resumo, o texto apresenta uma regulamentação orientada pelo risco e uma estrutura regulatória que se baseia nos direitos, visando a responsabilidade e a prestação de contas adequadas dos agentes econômicos que desenvolvem e usam a inteligência artificial. Ele também estabelece regras de responsabilidade civil, distinguindo entre sistemas de IA de alto risco ou excessivamente arriscados e sistemas de Inteligência Artificial de menor risco. Nos casos de sistemas de alto risco, os danos são atribuídos objetivamente aos fornecedores ou operadores, enquanto em sistemas de menor risco, a culpa do agente causador é presumida, favorecendo a vítima. Além disso, o texto define uma série de sanções administrativas que podem ser aplicadas às partes envolvidas em infrações às normas estabelecidas na Lei, abrangendo advertências, multas, publicização das infrações e até a suspensão das operações de sistemas de Inteligência Artificial, dependendo das circunstâncias e da gravidade das violações.

No tempo de desenvolvimento deste trabalho o Projeto de Lei 21/2020 se encontra no Senado, assim como o Projeto de Lei 2338 de 2023 e os Projetos de Lei nº 5.051, de 2019; 21, de 2020; 872, de 2021, tramitando em conjunto na Comissão Temporária Interna sobre Inteligência Artificial no Brasil, no Senado Federal, aguardando a emissão de relatório. Por sua vez o Projeto de Lei 4496 de 2019 se encontra na Comissão de Comunicação e Direito Digital, também no Senado Federal, aguardando audiência pública.

Já na União Europeia, por exemplo, havia a Diretiva de Proteção de Dados da União Europeia de 1995 (UNIÃO EUROPEIA, 1995) como um marco na legislação de privacidade na Europa. Foi adotada em 1995 e entrou em vigor em 1998, estabelecendo regras para a proteção de dados pessoais em todos os Estados membros da União Europeia. A diretiva foi criada com o objetivo de harmonizar as leis de privacidade de dados na Europa e, assim, garantir a livre circulação de dados pessoais dentro do mercado único europeu.

Essa Diretiva estabeleceu as bases para a proteção de dados pessoais na União Europeia e foi um passo importante para garantir a privacidade dos indivíduos na era digital. No entanto, com o avanço da tecnologia e a crescente quantidade de dados pessoais coletados e processados, a União Europeia percebeu a necessidade de atualizar sua legislação de proteção de dados, levando à adoção do Regulamento Geral de Proteção de Dados em 2016.

O Regulamento Geral de Proteção de Dados (UNIÃO EUROPEIA, 2016), por sua vez, foi aprovado em 2016 e entrou em vigor em maio de 2018, substituindo a Diretiva de Proteção de

Dados. O regulamento amplia e atualiza a proteção de dados pessoais, tornando as regras mais rigorosas e aplicáveis de forma consistente em todos os países membros da União Europeia.

Uma das principais diferenças entre a Diretiva e o regulamento está na natureza jurídica. A diretiva era um instrumento de harmonização que permitia aos Estados-membros implementarem a legislação de proteção de dados em seus próprios termos, enquanto o Regulamento Geral de Proteção de Dados é um regulamento diretamente aplicável a todos os Estados-membros da União Europeia, tornando a proteção de dados mais uniforme e consistente em todo o bloco.

Esse regulamento tem como objetivo a proteção de dados de pessoas naturais, tal como a lei brasileira e assim como ela, apresenta diretrizes para o tratamento de dados pessoais, definindo os atores que realizaram esse tratamento e estabelecendo direitos para o titular dos dados. Igualmente, o regulamento aborda o tema do tratamento automatizado de dados pessoais, em seu artigo 22, referindo-se às decisões individuais automatizadas. Desta forma, inclui-se o contexto da Inteligência Artificial neste regramento.

Contudo, no regulamento europeu, há um delineamento, em termos de limitação, mais apurado que a legislação brasileira quanto as decisões automatizadas. Isto porque o regulamento define apenas algumas hipóteses em que se permite esse tipo de tratamento. A primeira hipótese (artigo 22º, nº 2, alínea a) reside quando a decisão automatizada é necessária para celebrar ou executar um contrato entre o titular dos dados e um responsável pelo tratamento. Ou seja, se a tomada de decisão automatizada for essencial para a conclusão ou cumprimento de um contrato, o direito do titular dos dados de não ser submetido a essa decisão não será aplicado.

A segunda hipótese (artigo 22º, nº 2, alínea b) se refere a casos em que a decisão automatizada é autorizada pelo direito da União ou do Estado-Membro a que o responsável pelo tratamento estiver sujeito, desde que sejam previstas medidas adequadas para proteger os direitos e liberdades do titular dos dados. Isso significa que, se uma lei exigir uma decisão automatizada em determinado contexto, essa exigência prevalecerá sobre o direito do titular dos dados de não ser submetido a tal decisão. Já a terceira hipótese (artigo 22º, nº 2, alínea c) se aplica quando a decisão automatizada é baseada no consentimento explícito do titular dos dados. Nesse caso, o titular dos dados concorda explicitamente com a tomada de decisão automatizada, e portanto, não se aplicará o direito de não ser submetido a essa decisão.

Nos casos em que se têm a primeira e a terceira hipótese (artigo 22º, nº 2, alínea a e c) para tratamento automatizado, em tais circunstâncias, é necessário que o responsável pelo tratamento tome medidas adequadas para preservar os direitos, liberdades e interesses legítimos do titular dos dados, incluindo o direito de solicitar intervenção humana por parte do responsável, expressar sua opinião e contestar a decisão tomada, é o que se define o artigo 22º, nº 3.

Esses decisões automatizadas apontadas no artigo 22º do regulamento europeu, não podem fazer uso de dados de categorias especiais definidos no artigo 9º, nº 1. Isso quer dizer que não se pode tratar dados automaticamente, por Inteligência Artificial ou não, que se referem a origem racial ou étnica, de cunho político, sobre convicções religiosas, filosófica ou filiação

sindical, dados genéticos ou biométricos, de saúde, sobre a vida sexual ou orientação sexual do indivíduo, a menos que existe consentimento explícito do titular de dados ou se for necessário por motivos de interesse público importante.

O tratamento de dados é mais restritivo quando há automatização envolvida, o que inclui sistemas de Inteligência Artificial. A menos que sejam aplicadas as exceções mencionadas anteriormente, não é permitido o tratamento de dados especiais. Contar com o consentimento para esses tipos de tratamentos pode gerar problemas de gestão, uma vez que é necessário adquirir, manter e gerenciar os consentimentos fornecidos, o que pode se tornar complicado em sistemas inteligentes que requerem grandes quantidades de dados para funcionar. É comum que esses sistemas processem terabytes e até petabytes de dados.

Tanto a Lei Geral de Proteção de Dados quanto o Regulamento Geral de Proteção de Dados da União Europeia preveem a revisão de decisões automatizadas que afetam significativamente os direitos e liberdades dos titulares dos dados. No entanto, existem algumas diferenças entre as duas legislações. Uma das principais diferenças é que a Regulamento Geral de Proteção de Dados exige que as decisões automatizadas sejam revisadas por uma pessoa física, enquanto o Lei Geral de Proteção de Dados permite que essa revisão seja feita por meio de técnicas de inteligência artificial, desde que sejam implementadas salvaguardas adequadas para garantir a proteção dos direitos do titular dos dados.

Outra diferença é que a Lei Geral de Proteção de Dados permite que os titulares dos dados solicitem a revisão de decisões automatizadas que afetem seus interesses, mesmo que não sejam significativamente impactantes, enquanto o Regulamento Geral de Proteção de Dados exige que as decisões automatizadas sejam revisadas somente se produzirem efeitos significativos. Além disso, o Regulamento Geral de Proteção de Dados estabelece critérios específicos para a tomada de decisões automatizadas, incluindo a exigência de transparência e o direito do titular dos dados de ser informado sobre a existência de uma tomada de decisão automatizada, enquanto a Lei Geral de Proteção de Dados não possui disposições específicas sobre os critérios de tomada de decisão automatizada.

Por ser a legislação de proteção de dados brasileira mais permissiva se faz importante entender como se dá o uso da Inteligência Artificial no tratamento de dados pessoais, porque é um tema em destaque no contexto jurídico. Nesse sentido, é importante ressaltar que a proteção de dados é um direito fundamental e requisito essencial para qualquer tratamento de dados por Inteligência Artificial. No Brasil, a regulamentação da tecnologia tem relação direta com a proteção de dados, sendo essencial a compreensão das múltiplas relações jurídicas e da responsabilidade civil no tratamento de dados realizados por Inteligência Artificial.

Todavia, para que a utilização da Inteligência Artificial respeite os direitos e liberdades fundamentais do titular dos dados, é necessário que haja transparência e dever de informar significativamente sobre tratamentos automatizados. É nesse ponto que o direito à explicação de decisões automatizadas assume um papel crucial, pois é um fundamento para o exercício dos direitos dos titulares de dados pessoais, temas que serão abordados no Capítulo 4. Dessa forma,

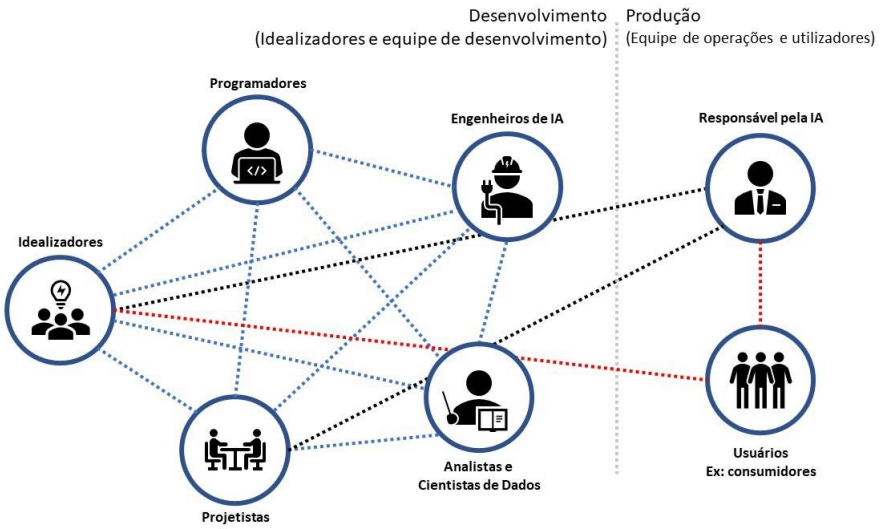
é essencial que se compreenda profunda do uso da Inteligência Artificial em relação à proteção de dados pessoais para garantir que os direitos fundamentais dos indivíduos sejam respeitados e que a ela seja utilizada de forma ética e responsável.

A responsabilidade civil sobre a Inteligência Artificial é outro ponto controverso e objeto de debate, no momento de confecção deste trabalho. A responsabilidade civil referente á Inteligência Artificial dentro de contexto da proteção de dados pessoais é discutida na Seção 3.3 deste capítulo. Evidenciar a responsabilidade em si detém relevância significativa no tocante a mostrar as consequências para aqueles que causem danos por conta da tecnologia em estudo. É igualmente necessário entender as fases e atores no ciclo de vida de um sistema de Inteligência Artificial para apontar de forma mais adequada essa responsabilidade.

3.3 As múltiplas relações jurídicas e a responsabilidade civil no tratamento de dados realizados por Inteligência Artificial

Inicialmente, é importante destacar que há uma cadeia de existência da Inteligência Artificial, onde diversos atores interagem, desde os idealizadores até os utilizadores. Os idealizadores são responsáveis pelos requisitos da atividade, enquanto os projetistas, normalmente uma equipe constituída por profissionais especializados, incluem engenheiros de Inteligência Artificial, analistas de sistemas, implementadores e cientistas de dados. Os utilizadores podem ser responsáveis pelo funcionamento e uso da Inteligência Artificial, como os funcionários de uma instituição, ou usuários externos, como consumidores que utilizam um serviço ou produto advindo de Inteligência Artificial. A Figura 9 mostra o relacionamento entre esses atores na fase de desenvolvimento e produção da Inteligência Artificial.

Figura 9 – Atores da cadeia de existência de Inteligência Artificial.



Fonte: O autor.

Na Figura 9, pode-se visualizar de forma simples a relação entre os atores dentro dessa

cadeia de existência da Inteligência Artificial. Na figura é possível observar diferentes relações específicas a partir dos traços tracejados em coloração azul, preta e vermelha. Os traços azuis representam as relações técnicas com atores especializados na tecnologia. Os traços pretos representam as relações internas de negócio com atores responsáveis por entenderem e operarem o negócio.

Os traços vermelho, por sua vez, representam as relações externas de negócio com atores que lidam direta ou indiretamente com usuários externos, como o exemplo trazido na figura, isto é, usuários consumidores. Embora exista uma divisão clara entre os ambientes de desenvolvimento e de produção, os idealizadores formam relações com todos os atores da cadeia de existência, não importando em que ambiente estejam. Isto porque são eles que, direta ou indiretamente, entregam a tecnologia para uso em ambiente real. A figura em questão mostra um rol exemplificativo dos atores e suas funções dentro de um projeto de desenvolvimento de software.

Exemplo disso está na figura do engenheiro jurídico, do qual é responsável por intermediar as necessidades do domínio jurídico e os projetistas que desenvolverão as tecnologias que serão utilizadas como ferramentas jurídicas por diversos profissionais do direito. Esse profissional em específico é capaz de realizar uma tradução da linguagem e lógica do direito para a da computação, facilitando as operações de modelagem e implementação dos programas de computador. Esses engenheiros jurídicos, em uma análise mais simplória, podem ser considerados como analistas especializados, do qual agregam ambos os conhecimento, o do direito e o da computação, para direcionar uma atividade específica como o de desenvolvimento de sistemas inteligentes de forma mais eficiente, rápida e fácil.

No texto do Projeto de Lei 2338 de 2023, é possível fazer um paralelo entre o que o projeto define como fornecedor de Inteligência Artificial - uma entidade que conceba um sistema de inteligência artificial, seja de forma direta ou mediante encomenda, visando à sua posterior disponibilização no mercado ou à sua utilização em um serviço fornecido por ela, sob a sua própria identificação ou marca, seja de forma remunerada ou gratuita - com os atores envolvidos no ambiente de desenvolvimento, isto é, os idealizadores e os projetistas e os operadores de Inteligência Artificial - uma entidade que faça uso, em benefício próprio, de um sistema de inteligência artificial, a menos que esse sistema esteja sendo empregado em uma atividade de natureza pessoal, sem caráter profissional - com os atores que estão em ambiente de produção, como a equipe de operações do sistema.

Na Seção 2.2, a cadeia de existência de um software, que também se aplicará a Inteligência Artificial, foi apresentada e agora será confrontada com as disposições contidas na Lei Geral de Proteção de Dados, no que diz respeito ao controlador e ao operador. Tanto o controlador quanto o operador podem desempenhar papéis em vários pontos dessa cadeia, seja no ambiente de desenvolvimento ou no ambiente de produção.

A Figura 2 da Seção 2.1 apresenta a cadeia de existência da Inteligência Artificial, pois sua forma é concretizada através de software, destacando as etapas de desenvolvimento,

produção, bem como os atores e atividades relacionados. No entanto, é importante destacar que a figura não contempla todos os atores envolvidos no âmbito da Lei Geral de Proteção de Dados.

Vale ressaltar que as atividades desempenhadas pelos projetistas, sempre em nome dos idealizadores, vão além do ambiente de desenvolvimento, seguindo as práticas comuns em projetos de software. Assim, os projetistas acompanharão o sistema, realizando manutenções para corrigir erros de funcionamento e segurança, além de aprimorar o software com a adição de novas funcionalidades e aprendizado a partir de novos dados coletados. É importante destacar que os atores apresentados na figura como projetistas são apenas exemplos, uma vez que a composição de equipes interdisciplinares pode requerer a presença de diversos profissionais.

No ambiente de desenvolvimento, o controlador pode conceber o projeto e determinar como o sistema deve funcionar para atender às expectativas da atividade final. Se possuir a expertise necessária, ele próprio pode implementar a tecnologia sem depender de outros atores. Além disso, o controlador pode utilizar a tecnologia desenvolvida diretamente em suas atividades ou negócios, saindo do ambiente de desenvolvimento e entrando no ambiente de produção. Já no ambiente de produção, o controlador pode atuar apenas como responsável pelo funcionamento da Inteligência Artificial em seu ambiente, não fazendo parte de nenhuma das fases de desenvolvimento.

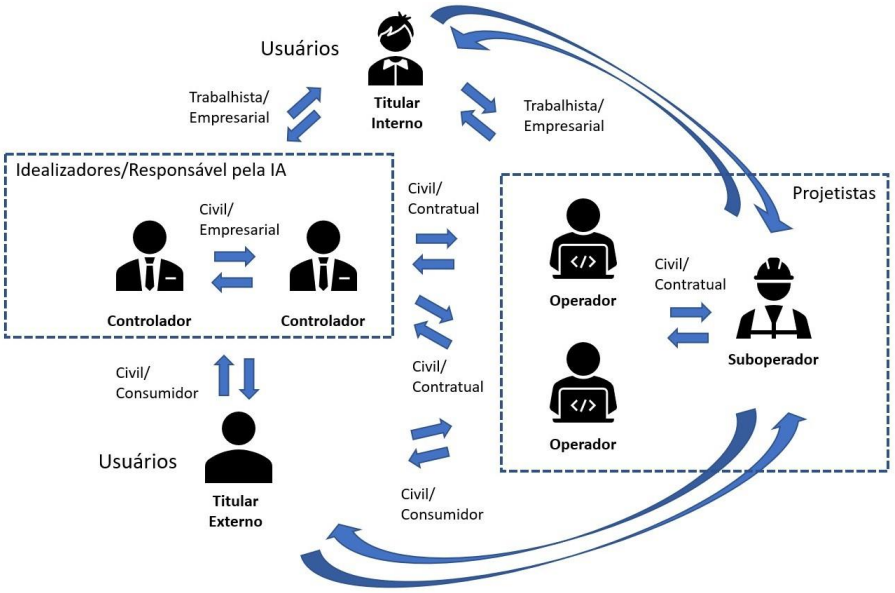
O responsável pela operação pode desempenhar o papel de projetista durante a fase de desenvolvimento, criando a tecnologia em nome do controlador. Nesse caso, ele pode continuar envolvido na manutenção e evolução da Inteligência Artificial no ambiente de produção, já que sistemas baseados em aprendizado de máquina exigem constante retroalimentação de dados para melhorar o desempenho. Durante a fase de desenvolvimento, o operador pode assumir diferentes atividades ou delegar algumas a terceiros especializados, como o "suboperador", que não é previsto em lei, mas foi definido pela Autoridade Nacional de Proteção de Dados (ANPD) (2022) como aquele que ajuda o operador a processar dados pessoais em nome do controlador. Dependendo da finalidade do sistema inteligente, pode ser necessário envolver vários operadores em atividades interdisciplinares.

Na fase de produção, o sistema de Inteligência Artificial será utilizado e acionado por usuários, que podem ser membros da equipe do controlador ou do operador (atuando em nome do controlador), bem como pelos próprios titulares de dados em situações de consumo ou outros cenários. Independentemente do caso, os titulares de dados serão os mais impactados, já que seus dados não só serão utilizados como fonte principal de funcionamento desses sistemas, como também serão o alvo dos resultados produzidos - os *outputs*.

Observa-se que existem diversas relações decorrentes do cenário apresentado na cadeia de existência da Inteligência Artificial, portanto, da mesma forma, haverá múltiplas relações jurídicas. É possível identificar relações entre controladores e operadores, operadores e suboperadores, controladores e funcionários, operadores e funcionários, colaboradores e titulares consumidores, operadores e titulares consumidores, dentre outras possíveis composições. Essas relações podem ser classificadas com base na cadeia de existência da

Inteligência Artificial, considerando as fases de desenvolvimento e produção. A Figura 10 montra as múltiplas relações existentes dentro da cadeia de existência da Inteligência Artificial, levando-se em consideração os papeis dos atores dentro do contexto da proteção de dados.

Figura 10 – Múltiplas relações na cadeia de existência de um sistema com Inteligência Artificial.



Fonte: O autor.

A Figura 10, portanto, ilustra as relações jurídicas entre os atores envolvidos na dinâmica de construção e funcionamento de sistemas de Inteligência Artificial. Essas relações incluem parcerias entre idealizadores e controladores responsáveis pela Inteligência Artificial, relações trabalhistas entre controladores/operadores e funcionários internos, contratos entre controladores e operadores, bem como relações entre operadores e suboperadores. Além disso, também existem relações de consumo entre controladores/operadores e titulares externos, entre outras.

Considerando essa cadeia complexa, com vários atores desempenhando diferentes papéis, muitas vezes confundíveis, e múltiplas relações jurídicas envolvidas, a análise da responsabilidade civil se torna igualmente complicada. Isso é especialmente verdadeiro para sistemas de Inteligência Artificial com aprendizado de máquina ou aprendizado profundo, que não oferecem, por padrão de desenvolvimento, a garantia de resultados precisos e, dependendo do modelo utilizado, nem mesmo uma explicação de como o processo ocorreu.

Deve-se destacar os riscos associados ao uso dessa tecnologia digital, devido à incerteza dos resultados que ela pode produzir. Mesmo que todo o processo de desenvolvimento tenha sido conduzido com cuidado e precaução, levando em consideração todos os pontos necessários para preservar a privacidade dos dados, com o modelo de Inteligência Artificial testado e validado, assim como os dados utilizados para treinamento e testes “saudáveis” e sem vieses discriminatórios ou que resultem em violação de direitos, no ambiente de produção, os dados que alimentam continuamente o sistema de aprendizado podem alterar o funcionamento do

sistema para um estado enviesado com resultados indesejáveis, caso não haja controle e auditoria adequados.

Um software inteligente com deficiências na qualidade devido à sua falta de segurança não pode ser aceito em ambientes de produção ou no mercado, e é dever do fornecedor criador intervir para resolver, reduzir ou eliminar os riscos associados ao produto de Inteligência Artificial fornecido à sociedade. Isso estabelece uma obrigação para o fornecedor criador de identificar os responsáveis pelo uso da tecnologia, a fim de corrigir as deficiências. O *Recall* é uma das maneiras de cumprir essa obrigação (ALBUQUERQUE; JÚNIOR, 2018). É responsabilidade do usuário da Inteligência Artificial, no ambiente de produção, seguir as instruções do fornecedor criador para corrigir as deficiências. Se o usuário da Inteligência Artificial não cumprir o *Recall*, ele terá uma responsabilidade maior em caso de danos decorrentes do uso do sistema.

Entretanto, dentro do contexto de proteção de dados pessoais, não há indicação clara do regime de responsabilidade civil adotado. A literatura especializada tenta sugerir qual regime pode ser aplicado, mas não há um consenso claro sobre isso, nem um posicionamento consolidado nos tribunais brasileiros até o momento. Por essa razão, é necessário analisar inicialmente os pontos em que a responsabilidade civil é mencionada na legislação específica. Na Lei Geral de Proteção de Dados Pessoais, mais especificamente na Seção III do Capítulo VI, que engloba os artigos 42 a 45, há uma abordagem explícita sobre a responsabilidade civil e o ressarcimento de dados.

Os parágrafos do artigo 42 da Lei Geral de Proteção de Dados estabelecem diretrizes para a atribuição de responsabilidade, distribuindo-a entre controladores e operadores, e indicando a possibilidade de inversão do ônus da prova durante processos judiciais, bem como o direito de regresso. Por conseguinte, os aspectos mais relevantes para a responsabilidade civil estão contidos no caput do artigo 42 e nos artigos 43, 44 e 45. O caput do artigo 42 prevê a alocação de responsabilidade aos agentes de tratamento em casos de violação da legislação de proteção de dados, obrigando-os a reparar os danos decorrentes de suas atividades de tratamento

No que diz respeito ao artigo 43 da lei, são apresentadas as exclusões de responsabilidade que podem ser invocadas pelos agentes de tratamento em casos nos quais não realizaram o tratamento em si, ou que, ainda que tenham realizado algum tratamento, não houve violação da legislação de proteção de dados, ou ainda, em que o dano tenha sido causado por culpa de terceiros ou dos próprios titulares dos dados. O artigo 44, por sua vez, define a responsabilidade por tratamento inadequado de dados quando os agentes de tratamento deixarem de cumprir a legislação ou não fornecerem a segurança esperada para realizar a atividade de tratamento.

O parágrafo único do mesmo artigo obriga os agentes de tratamento a observarem o disposto no artigo 46, que trata dos controles de segurança de dados que previnem situações acidentais e ilícitas que possam ocasionar tratamento inadequado. Alguns aspectos circunstanciais que afetam a análise de tratamento inadequado são relevantes e apresentados nos incisos I, II e III do artigo 44: o modo pelo qual o tratamento é realizado em um caso concreto,

a resultante desse tratamento e os riscos esperados, e o estado atual da tecnologia utilizada na época do tratamento. Já o artigo 45 direciona a responsabilidade pelo tratamento de dados no âmbito das relações de consumo para a legislação específica, ou seja, o Código de Defesa do Consumidor é tomado como referência.

Em resumo, a Lei Geral de Proteção de Dados Pessoais apresenta diretrizes sobre a atribuição de responsabilidade civil em casos de violação da legislação de proteção de dados. Contudo, o uso de sistemas inteligentes baseados em Inteligência Artificial apresenta desafios adicionais para a determinação do regime de responsabilidade civil aplicável. É importante analisar as circunstâncias específicas de cada caso, considerando fatores como o nível de autonomia e o fator de interação humana dentro do processo do sistema.

Esse modo agressivo de tratamento é uma circunstância relevante para avaliar se o responsável pela Inteligência Artificial está fornecendo a segurança necessária ao titular dos dados, em conformidade com o inciso I do artigo 44 da Lei Geral de Proteção de Dados. É incontestável que a utilização de sistemas inteligentes baseados em Inteligência Artificial implica na obrigação de reparar os danos decorrentes de tratamento inadequado ou ilegal. No entanto, o desafio neste contexto é determinar qual regime de responsabilidade civil deve ser aplicado. Isso ocorre porque o tratamento realizado por Inteligência Artificial envolve a automação em grande escala de processos que frequentemente classificam pessoas em perfis para ajudar na tomada de decisões. Isso elimina o elemento humano do processo de tomada de decisão deixando-o como mero monitor/observador, ampliando a quantidade e o alcance do tratamento.

Uma prática arriscada é utilizar a aprendizagem de máquina por meio de modelos de Inteligência Artificial que não possuem uma explicação satisfatória sobre seu funcionamento, especialmente quando não se pode garantir um resultado esperado (LUGER, 2013). Esse cenário representa um nível de risco significativo, o que é mais um ponto relevante para avaliar a segurança, conforme destacado no inciso II do artigo 44. Atualmente, não existe um padrão de explicabilidade que permita avaliar com precisão a acurácia dos modelos de aprendizagem de máquina e aprendizagem profunda, o que deve ser considerado na verificação das circunstâncias descritas no inciso III do mesmo artigo.

Compreender a cadeia de existência da Inteligência Artificial e os diversos atores envolvidos é fundamental para uma análise precisa da responsabilidade civil, já que cada um desses atores tem diferentes papéis e funções.

Conforme explicado no início da atual seção, a Lei Geral de Proteção de Dados não traz uma definição explícita sobre o regime de responsabilidade civil. No entanto, a técnica legislativa adotada no Brasil, em casos como este em que não há menção a expressões como “independentemente de culpa” ou “responde objetivamente”, geralmente adota o regime subjetivo de responsabilidade civil como regra.

Na Lei Geral de Proteção de Dados, são estabelecidos deveres para os responsáveis pelo tratamento de dados, cujo não cumprimento resultaria em violação da legislação e tornaria o

tratamento irregular. Esses mesmos deveres indicam uma preocupação do legislador em deixar claro que eles não seriam atribuídos caso o regime de responsabilidade civil fosse diferente do subjetivo.

Assim, como não há sistema totalmente seguro devido à inviabilidade técnica de alcançá-lo, mesmo na ocorrência de dano e nexo de causalidade, se os responsáveis pelo tratamento utilizarem todas as medidas possíveis para evitar prejuízos, não deveria haver responsabilidade e a verificação estaria no contexto da culpa. O substitutivo do Projeto de Lei nº 21/2020 reafirma a intenção do legislador em regular a utilização de Inteligência Artificial no Brasil. O artigo 6º, inciso VI, propõe explicitamente a atribuição de responsabilidade subjetiva aos atores envolvidos na cadeia de existência da Inteligência Artificial, de acordo com sua participação. Contudo, o Projeto de Lei 2338 de 2023 traz em seu contexto de responsabilidade, uma modelagem baseada em riscos, estabelecendo para os agentes de Inteligência Artificial que desenvolvem e operam sistemas inteligentes considerados de risco excessivo e alto, a responsabilidade objetiva, assemelhando-se com parte significativa da literatura que defende que esse regime seja predominante ao se tratar de Inteligência Artificial. No entanto, para sistemas que não estejam classificados como de alto risco, a responsabilidade ainda contém o elemento de culpa, porém em sua presunção, invertendo o ônus da prova em desfavor aos agentes citados.

Dado o contexto complexo das relações estabelecidas no uso de Inteligência Artificial e dos papéis claramente definidos para cada ator envolvido, a culpa se torna um requisito fundamental para a responsabilidade civil. Cada ator terá que demonstrar que não agiu com negligência, imprudência ou imperícia e que cumpriu adequadamente as normas aplicáveis e o sistema de proteção de dados pessoais dentro deste cenário.

Aqueles que defendem o regime subjetivo[1] de responsabilidade civil estão preocupados com o possível impacto negativo no avanço da tecnologia, especialmente no caso da Inteligência Artificial, se a responsabilidade civil for excessivamente restritiva. Isso poderia inibir seu uso, desacelerando todo o ecossistema de inovação que depende dessa tecnologia disruptiva, com impacto significativo nos setores econômico, social, educacional e em vários outros setores importantes. Se a responsabilidade civil for subjetiva, a atividade teria menos risco, incentivando o desenvolvimento da tecnologia. Por outro lado, se houver responsabilidade objetiva, o risco da atividade seria maior, o que reduziria esse incentivo.

A adoção do regime subjetivo de responsabilidade civil como direcionador do sistema com Inteligência Artificial implicaria na necessidade de analisar sempre a existência de culpa para todos os envolvidos na cadeia de existência da Inteligência Artificial. Diante do potencial prejuízo que esses sistemas podem causar, a balança sempre tenderia a favorecer os controladores, operadores, suboperadores e responsáveis pela Inteligência Artificial, em detrimento do titular dos dados pessoais, que possui um atributo semelhante ao do consumidor: a vulnerabilidade. Essa vulnerabilidade é intensificada pelo uso de Inteligência Artificial com

[1]São expoentes desta corrente GUEDES (2019) e Bioni e Dias (2020), que baseiam seus argumentos em uma responsabilidade civil com um grau significativo de objetividade.

técnicas e modelos que não permitem uma explicação satisfatória do tratamento automatizado.

Um fator importante é a possibilidade de viés no sistema de Inteligência Artificial devido aos dados consumidos em ambiente de produção, pois isso pode alterar seu comportamento. Assim, a atividade de monitoramento e saneamento constante dos dados é crucial para o funcionamento adequado. Alguns aspectos da responsabilidade subjetiva devem ser reconhecidos, já que a dinâmica das relações exige a identificação de quanto cada um detém de responsabilidade e se cumpriu cuidadosamente suas atividades.

A posição majoritária na literatura especializada defende a responsabilidade objetiva[2], que se baseia na interpretação sistemática da Lei Geral de Proteção de Dados e do Código Civil. Essa interpretação se baseia principalmente no parágrafo único do artigo 927 do Código Civil de 2002 (BRASIL, 2002), que aponta para o risco da atividade. Para as relações de consumo envolvendo titulares de dados e fornecedores de produtos e serviços por meio de Inteligência Artificial, a responsabilidade objetiva é regulada pelo Código de Defesa do Consumidor e pelo artigo 45 da Lei Geral de Proteção de Dados. Para o poder público, o tratamento automatizado por Inteligência Artificial também atrai a responsabilidade objetiva, como estabelecido no artigo 37, §6º, da Constituição de 1988.

Uma falha nas justificativas de tratamento de dados pessoais pode indicar um risco inerente ao processamento desses dados, e em situações em que ocorrem danos significativos afetando direitos difusos (TAMBOSI, 2021), o titular dos dados pode se encontrar em uma posição vulnerável em relação aos responsáveis pelo projeto e idealização da tecnologia. Nessas circunstâncias, a culpa pode ser irrelevante para a obrigação de reparar o dano, tornando uma abordagem de responsabilidade civil objetiva mais apropriada. Afinal, os projetistas e idealizadores da tecnologia possuem mais recursos e capacidade para prevenir possíveis danos do que o titular dos dados, que não deve ser responsabilizado pelos prejuízos decorrentes do uso da tecnologia.

Uma outra indicação pode ser encontrada no artigo 43 da Lei Geral de Proteção de Dados, que trata das exclusões de responsabilidade e apresenta semelhanças com as exclusões do sistema de consumo. Uma vez que a lei traz apenas hipóteses restritas de exclusão de responsabilidade, isso sugere uma inclinação para adotar o regime de responsabilidade objetiva. Ao se adotar estritamente a responsabilidade objetiva para os controladores, operadores e agentes de Inteligência Artificial, é ignorada a parcela de responsabilidade de cada um deles, e não é relevante observar os deveres impostos pela legislação, como já mencionado nesta seção. Mesmo que um determinado ator tenha desempenhado sua atividade de forma perfeita, ele será responsabilizado por falhas de outros que não o fizeram.

Entretanto, estabelecer uma única forma de responsabilidade em uma estrutura tão complexa como a da produção de Inteligência Artificial, sem levar em conta a relação jurídica em questão, pode gerar instabilidade e insegurança jurídica. A adoção de um regime de responsabilidade civil múltiplo pode ser uma solução para esse problema. O primeiro passo seria

[2]Defendem esta visão Mulholland (2020), Gondim (2021), com base no risco da atividade, principalmente.

identificar a relação jurídica envolvida e, a partir disso, definir a responsabilidade e a medida para cada um dos envolvidos. Essa investigação seria feita caso a caso. Além disso, o grau de autonomia do sistema também precisa ser avaliado, já que quanto maior o nível de autonomia, maior o risco envolvido. Nesse sentido, a transição de uma responsabilidade subjetiva para uma objetiva parece ser razoável.

O uso de múltiplos regimes de responsabilidade também é evidente na União Europeia, conforme explicado por MEDON (2022). O Regulamento do Parlamento Europeu de 20 de outubro de 2020 sugere a adoção de um sistema tripartite composto por: 1) regime de produtos defeituosos; 2) regime dual, dependendo do tipo de Inteligência Artificial, já que pode haver sistemas de baixo e alto risco; e 3) regime subjetivo, como último recurso, considerando a legislação dos países membros. O autor destaca que ainda existe a responsabilidade contratual.

A cadeia de produção da Inteligência Artificial envolve múltiplos atores, o que resulta em várias relações jurídicas que precisam ser consideradas na análise das causas de danos decorrentes do uso da tecnologia. Esses danos podem ser causados por diversos fatores, como mau funcionamento do próprio modelo, uso enviesado ou viciado de dados utilizados na produção, ou requisitos de sistemas que impõem um funcionamento direcionado. Além disso, esses fatores podem ocorrer simultaneamente, não havendo uma única causa para o dano. Portanto, a causalidade alternativa[3] é uma possibilidade para lidar com a incerteza sobre a causa quando há duas ou mais potenciais causas entre as várias causalidades múltiplas.

[3]A causalidade alternativa ocorre quando não há certeza sobre qual causa específica deu origem a determinado resultado (SANTOLIM, 2014).

4 DO TRATAMENTO AUTOMATIZADO E O DIREITO À EXPLICAÇÃO NO CONTEXTO DE INTELIGÊNCIA ARTIFICIAL

Este capítulo tem como principal objetivo analisar como a transparência se apresenta em seus níveis e como ela pode tomar forma para substanciar uma explicação suficientemente significativa com o propósito prover o titular de dados pessoas com informações relevantes para entender um tratamento automatizado, incluindo-se por sistemas com Inteligência Artificial. Para isso, primeiramente discute-se a relação do dever de informar sobre o tratamento de dados pessoais e sua relação com os níveis de transparência. Isto permitirá entender a transparência por suas características e sugerir quais níveis se sobressaem em termos de adequação com a legislação.

Após, se apresentará a discussão sobre a explicação significativa e como o direito à explicação de decisões automatizadas, previsto na Lei Geral de Proteção de Dados, pode servir como base para que o titular de dados possa exercer outros direitos inerentes de sua condição. Isso permitirá a compreensão de como uma explicação significativa poderá ser construída e como ela pode ser aplicada em diversos cenários de aplicação, incluindo-se dentro de processos judiciais, conforme discutido no Capítulo 5, como forma de concretizar o princípio da motivação das decisões judiciais quando estas são proferidas com auxílio de sistemas inteligentes.

Por essa razão, no capítulo anterior, discutiu-se os impactos da Inteligência Artificial nos direitos fundamentais, assim como as tentativas de regulamentação no Brasil. Isso ocorreu devido a importância dos possíveis riscos que esses sistemas podem trazer, violando direitos com decisões desumanizadas e com potencial de afetar o indivíduo e a sociedade de maneira imprevisível. Também foi abordada a responsabilização civil para aqueles que não observam os direitos e garantias trazidos pela Constituição e pela legislação que trata do tema, assim como as consequências para os atores que desenvolveram ou desenvolverão a tecnologia e para aqueles que a utilizam em suas atividades de negócio.

Com esse propósito, a metodologia utilizada conta com uma revisão de literatura narrativa focada nos objetivos deste capítulo. Buscou-se, então, por publicações científicas como artigos científicos nacionais e internacionais, incluindo-se literatura cinzenta, conforme os protocolos nos Apêndices A e B.

4.1 Os níveis de transparência e o dever de informar significativamente sobre tratamentos automatizados

A transparência pode ser definida como a qualidade ou estado de ser aberto, claro e honesto sobre as informações que são compartilhadas. Em outras palavras, a transparência refere-se à prática de fornecer informações completas e precisas sobre algo, sem nada ocultar. A transparência parece exercer um papel de vital importância dentro da sociedade, assim como no sistema jurídico normativo brasileiro. Com ela é possível exercer fiscalização nas atividades estatais, ter segurança jurídica no cumprimentos de contratos, saber como os dados pessoais

estão sendo tratados, saber como um determinado sistema de software com Inteligência Artificial funciona e quais os impactos para quem se submete a esse sistema, etc.

No contexto da transparência, a opacidade é uma característica de singular significado dado que define o grau de percepção e visualização do que se quer transmitir. Esse ponto em específico determina o quanto será exposto a um terceiro observador e assim como na física, onde a transparência pode ser medida e avaliada por meio de parâmetros como índice de refração, absorção de luz, transmitância, entre outros, no contexto informacional e de proteção de dados ocorre o mesmo, ou seja, é possível estabelecer critérios para avaliar o nível de transparência aplicado (FELZMANN et al., 2019).

Assim, o nível de transparência depende do nível de opacidade atribuído ao que se quer expor. Por esta razão, é possível operar a transparência em diferentes níveis de granulação, são eles: 1) transparência total, onde não há ocultação, implicando na divulgação completa de todos os dados e informações disponíveis; 2) transparência parcial, alguns dos dados e informações são retidos por motivos e interesses legítimos, a exemplo da privacidade ou confidencialidade. No entanto, a maioria das informação são abertamente compartilhadas; 3) transparência seletiva, onde existe apenas o compartilhamento de informações específicas selecionadas por razões específicas; e 4) transparência limitada, envolve a divulgação de uma quantidade mínima de informações necessárias para comprimento de obrigação legal ou regulatória (FELZMANN et al., 2019; BLACKLAWS, 2018; CRUZ et al., 2016; NUNES; CAPPELLI; RALHA, 2017).

O Tabela 1 sumariza os níveis de transparência comparando-os com base em suas principais características para melhor compreensão. Como características foram selecionadas: o acesso, a abrangência, a divulgação, a clareza, o objetivo da transparência, o grau de transparência e a confiança que cada nível fornece. Com a análise dessas características é possível sugerir quais níveis podem ser os mais adequados dentro do contexto da Lei Geral de Proteção de Dados ao se tratar do direito à explicação de decisões automatizadas.

Tabela 1 – Comparativo entre níveis de transparência.

Características	**Transparência Total**	**Transparência Parcial**	**Transparência Seletiva**	**Transparência Limitada**
Acesso	Acesso irrestrito a todas as informações	Acesso restrito a algumas informações	Acesso concedido a determinados grupos ou pessoas	Acesso limitado a informações específicas
Abrangência	Todas as informações disponíveis	Algumas informações disponíveis	Informações disponíveis para um grupo seleto	Informações disponíveis apenas para fins específicos

Divulgação	Divulgação proativa de todas as informações	Divulgação seletiva de informações	Divulgação controlada de informações	Divulgação restrita de informações
Clareza	Informações apresentadas de forma clara e compreensível	Informações apresentadas de forma parcial ou confusa	Informações apresentadas de forma seletiva	Informações apresentadas de forma limitada e específica
Objetivo da Transparência	Promover a confiança e a prestação de contas	Promover a confiança e a prestação de contas em áreas específicas	Proteger informações sensíveis	Proteger informações confidenciais
Grau de Transparência	Alta transparência em todas as áreas	Grau variável de transparência em áreas específicas	Baixo grau de transparência em áreas específicas	Baixo grau de transparência em todas as áreas
Confiança	Confiança máxima na divulgação das informações	Confiança limitada na divulgação das informações	Confiança limitada na seleção das informações	Pouca confiança na divulgação das informações

Fonte: Adaptada de Nunes, Cappelli e Ralha (2017), Felzmann et al. (2019), Blacklaws (2018) e Cruz et al. (2016).

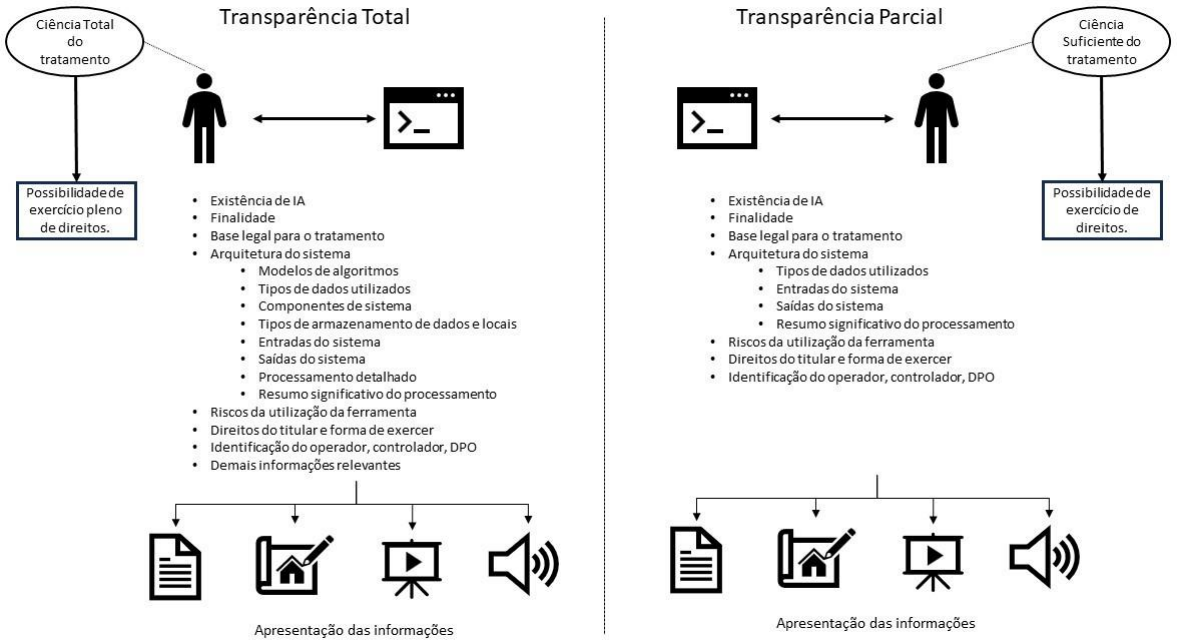
O Tabela 1 mostra que a transparência total elimina o problema da opacidade por completo, trazendo mais confiabilidade nos processos em que é aplicada. Por outro lado, a transparência limitada configura-se como seu oposto, trazendo um grau elevado de opacidade e dificultando a compreensão dos processos que o adota.

A transparência total envolve o acesso irrestrito e claro a informações não tendenciosas e abrangentes, atualizadas regularmente. Seu propósito é fortalecer a confiança, credibilidade e responsabilidade entre partes, sejam organizações, governos ou indivíduos. Isso facilita a tomada de decisões informadas e medidas apropriadas, reduzindo a corrupção e promovendo a confiança nas instituições. (LARSSON; HEINTZ, 2020). Na prática, o usuário teria a ciência exata e completa sobre o funcionamento da ferramenta inteligente, a começar da informação da existência de uso da ferramenta, seus parâmetros, finalidade e demais informações necessárias. Isso proporciona um controle maior do usuário titular de dados, inclusive para o fornecimento de seu consentimento, caso esta seja a base legal para o tratamento. Essa comunicação pode ser construída utilizando recursos multimídia, com o objetivo de alcançar todos os níveis de usuário.

A transparência parcial envolve reter parte das informações devido a limitações legítimas,

acrescentando restrições ao acesso. Suas características incluem divulgação seletiva, acesso restrito, clareza, honestidade e abrangência direcionada. Esse nível busca equilibrar a divulgação de informações com a proteção de dados sensíveis, sendo relevante em áreas como segurança nacional e finanças pessoais. O objetivo é aumentar a confiança pública, desde que a restrição seja justificada e aplicada somente a informações legítimas e relevantes. (BAROCAS; HARDT; NARAYANAN, 2017). Na prática, o titular de dados teria ciência de informações relevantes e significativas sobre todo o processo, como as entradas e saídas do sistema, explicação simplificada do processamento, armazenamento e comunicação dos dados pelo sistema, mas os detalhes técnicos e demais informações de negócio não seriam expostos ao usuário. Neste caso, a verificação das informações só poderia ser executada observando a descrição do procedimento e comparando as entradas com possíveis saídas esperadas. Aqui, a explicação do processamento se torna significativa. A Figura 11 mostra um comparativo entre os níveis total e parcial de transparência.

Figura 11 – Transparência total e parcial.



Fonte: O autor.

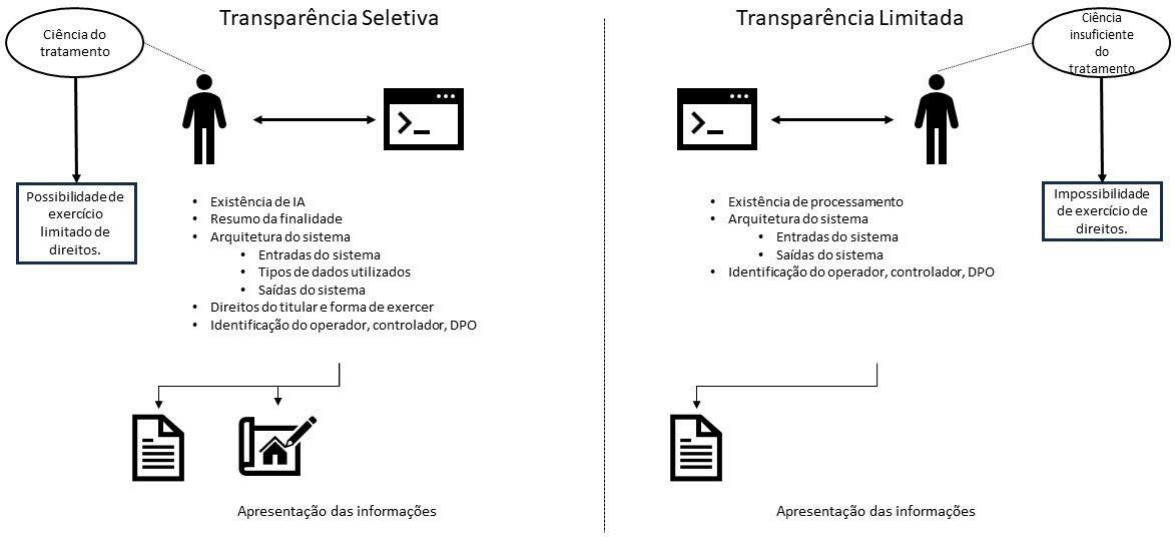
A transparência seletiva implica na restrição significativa da exposição de dados e informações, permitindo a manipulação e a apresentação tendenciosa para atender a interesses específicos. Suas características incluem a divulgação seletiva para controlar a percepção do receptor, o acesso restrito direcionado a grupos selecionados, a opacidade na apresentação dos dados, a desonestidade com informações tendenciosas e a falta de responsabilidade ao se recusar a assumir ações corretivas por divulgação seletiva. Seu propósito é controlar a percepção pública, proteger interesses particulares à custa da verdade e integridade, frequentemente

por governos, empresas ou indivíduos que buscam manipular a opinião pública ou ocultar informações desfavoráveis (WACHTER; MITTELSTADT; FLORIDI, 2017a).

Esse tipo de transparência é prejudicial à sociedade, pois pode impedir decisões informadas e minar a confiança nas instituições, sendo considerada antiética e sujeita a sérias consequências para os envolvidos. O usuário, como titular de dados, teria acesso à poucas informações relevantes. Seria informado apenas sobre a existência de uma ferramenta automatizada para processamento dos dados, com resumo de finalidade e o que das entradas resultariam as saídas do sistema.

A transparência limitada é o nível mais baixo de transparência em que algumas informações são divulgadas, mas de forma restrita ou insuficiente. Suas características incluem a divulgação escassa e incompleta de poucos dados, o acesso restrito, a falta de clareza com informações técnicas ou ambíguas, a tendenciosidade na manipulação da percepção do receptor e a falta de abrangência, retirando informações relevantes de propósito. Seu objetivo é controlar a percepção do público, muitas vezes para proteger informações sensíveis ou interesses específicos, mas prejudica a confiança nas instituições e impede decisões informadas. A divulgação completa e honesta das informações é crucial para manter a confiança e a integridade das instituições (WACHTER; MITTELSTADT; FLORIDI, 2017a). Aqui o usuário seria informado que haverá um processamento, sem especificação de automatização ou humanização e conhecerá as entradas do sistema, pois normalmente ele fornecerá tais dados, e saídas do sistema como resultado. Não há qualquer complemento de informação para o usuário, que ficará a mercê da plataforma.

Figura 12 – Transparência seletiva e limitada.



Fonte: O autor.

A Figura 12 apresenta um comparativo entre os níveis de transparência seletiva e limitada. É possível também, observando a Figura 11, fazer um comparativo entre os níveis

de transparência apresentados.

A escolha do nível de transparência dependerá do contexto e da finalidade da comunicação. Em geral, quanto mais aberto e claro for o compartilhamento de dados e informações, maior será a transparência e a confiança que podem ser estabelecidas entre as partes envolvidas. Ressalta-se que quando se envolve o tratamento de dados realizados por sistema de Inteligência Artificial deve-se atentar para um nível de transparência condizente com o esperado e disciplinado pela legislação específica.

Isso porque, levando-se em consideração o microssistema de proteção de dados pessoais no Brasil, em especial a Lei Geral de Proteção de Dados, em seu artigo 6º, inciso VI, o qual define o princípio da transparência, o nível de transparência oferecido ao titular de dados precisa se adequar, no mínimo, ao disposto como exigência legal, isto é, informações claras e precisas sobre o fora coletado e processado. Assim, esse princípio é essencial para garantir que as organizações tratem os dados pessoais dos titulares com respeito e responsabilidade. Ao tornar as informações claras e acessíveis, a Lei Geral de Proteção de Dados busca proteger a privacidade dos indivíduos, fornecendo transparência em todas as etapas do tratamento dos dados pessoais (MONTEIRO, 2018).

Contudo, há ao final do mesmo inciso a expressão "observados os segredos comercial e industrial". Isso indica uma certa restrição à transparência em termos de exposição do tipo de conteúdo que deve ser divulgado. Existe, então, uma permissão legal para omissão ou restrição dos dados e das informações que serão disponibilizadas ao titular de dados por conta de interesse legítimo do controlador e operador, pois se assim não fosse, eles seriam obrigados legalmente a fornecer segredos de sua atividade de negócio, o que poderia causar danos irreparáveis ao negócio.

Na esfera do consumo, por exemplo, o nível de transparência total seria o mais adequado para a proteção do consumidor, pois há o dever de informação atribuído ao fornecedor do produto ou serviço que se utiliza de sistemas inteligentes, levando-se em consideração o estado de constante vulnerabilidade do consumidor. Esse nível, portanto, cumpriria os requisitos de transparências exigidos perante a legislação de consumo, do qual deriva do direito do consumidor em se ter informações claras e transparentes do que concerne à especificação correta de quantidade, características, composição, qualidade e, sobretudo os riscos inerentes da utilização do sistema, trazido no inciso III do artigo 6º do Código de Defesa do Consumidor.

O Código de Defesa do Consumidor se torna relevante quanto à Inteligência Artificial, uma vez que os sistemas inteligentes são postos em produção em cenário de consumo, criando perfil de consumo, direcionando comportamento do consumidor e limitando, inclusive, os seus passos e seu alcance. Neste sentido, o Código de Defesa do Consumidor estabelece que as interações entre consumidores e fornecedores devem ser caracterizadas pela transparência. Isso significa que ambas as partes têm a responsabilidade de agir com honestidade e sinceridade antes, durante e após a negociação. Nesse contexto, as informações fornecidas ao consumidor são consideradas parte integrante do contrato, e quaisquer cláusulas que possam restringir os direitos

do consumidor devem ser redigidas de forma clara e destacada, de modo que o consumidor possa compreendê-las imediatamente.

Nos casos em que se utilizam plataformas com Inteligência Artificial, os termos de uso, assim como os avisos de privacidade e quaisquer outros mecanismos, a exemplo dos *cookies* para identificação de usuário e restauração de sessão de comunicação entre a plataforma e o usuário consumidor/titular de dados, que fazem referência de alguma forma aos sistemas disponíveis precisam implementar instrumentos de transparência, no sentido de prover ao consumidor todas as informações necessárias para o consumo da ferramenta com Inteligência Artificial ou através dela. Por consequência, as informações e notificações ao consumidor/usuário/titular de dados sobre o tratamento de seus dados por um sistema inteligente deve expor todos os processos que o impactam, mantendo, assim, a conformidade com o dever de informação imposto ao fornecedor/controlador/operador no contexto de consumo e proteção de dados.

Há, desta forma, conexão entre o princípio da transparência na esfera do consumo com o princípio da transparência vindo na Lei Geral de Proteção de Dados. Neste sentido, em ambos os casos, a intenção reside na proteção da parte hipossuficiente da relação, ou seja, a vulnerabilidade que tanto o consumidor quanto o titular de dados possuem. Em muitas relações de consumo, a figura do consumidor restará unificada com a do titular de dados pessoais, fazendo que com que o microssistema de proteção de dados se relacione intimamente com o microssistema consumerista.

Além disso, o princípio da transparência, trazidos pela Lei Geral de Proteção de Dados, dialoga com os demais princípios incorporados na mesma lei no sentido de englobá-los quanto de sua concretização. Espera-se que a finalidade de um tratamento seja exposta, assim como a necessidade do tratamento e a adequação entre este tratamento necessário e a finalidade sejam divulgados de maneira clara e precisa, como preconiza o princípio do livre acesso. Além disso, a transparência direciona a compartilhamento das informações como as medidas de segurança e as medidas preventivas adotadas para determinado tratamento. A transparência também é a base para a responsabilização e para a prestação de contas, afinal, os controladores são responsáveis pelo tratamento e devem prestar contas às autoridades competentes e para os titulares de dados.

De mesma forma, há uma relação intrínseca entre a transparência e o direito do titular de dados de ser informado sobre determinado tratamento do qual é submetido. Esse direito está previsto no artigo 9ª da Lei Geral de Proteção de Dados e indica que o controlador de dados deve informar ao titular, de maneira específica, facilitada, completa e clara sobre a finalidade, forma e duração do tratamento, a identificação do controlador e as responsabilidades dos agentes que realizaram o tratamento, dentre outras informações relevantes para o titular. Isso inclui informações suficientes para que o titular de dados possa fornecer seu consentimento para o tratamento, principalmente quando se utiliza de tratamento automatizado. Essas informações necessitam de destaque, além da uma explicação do impacto da ferramenta para o titular de dados, ou seja, todos os critérios de uma explicação *ex ante* (Seção 4.2) precisa ser informado.

O direito de ser informado é uma garantia importante para os titulares de dados pessoais,

uma vez que permite que eles tenham um grau de controle maior sobre o tratamento de seus dados e possam tomar decisões informadas sobre a divulgação de seus dados pessoais. A não observação desse direito pode resultar em sanções para os agentes de dados que não respeitam a Lei Geral de Proteção de Dados. Por conseguinte, o princípio de transparência fornece a base para a concretização deste direito.

O mesmo ocorre na esfera do consumidor, o direito de ser informado contidos em vários artigos do Código de Defesa do Consumidor, incluindo o 4º, 6º (inciso III), 8º, 31, 37 (§3º), 46 e 54 (§3º e §4º) também se consubstancia no princípio da transparência. Esse direito igualmente objetiva garantir que o consumidor tenha completa ciência da natureza e alcance exatos das responsabilidades que está assumindo em relação ao fornecedor e quais responsabilidades o próprio fornecedor assume em relação ao consumidor.

Ainda quanto ao direito de ser informado, o titular de dados precisa ter ciência sobre a existência de incidente de segurança da dados que envolvam seu dados. Os controladores e operadores devem divulgar tais acidentes, o que permite, em princípio, que os titulares da dados possam tomar as medidas cabíveis para se proteger. O mesmo ocorre com os parâmetros voltados às políticas de privacidade e de proteção de dados que circundam todas as atividades e processos do controlador e do operador no que couber. A visibilidade proporcionada pela transparência garante que os titulares não fiquem a mercê de qualquer tipo de tratamento de dados (WACHTER; MITTELSTADT; RUSSELL, 2017). Na relação de consumo o mesmo cuidado deve ser observado perante ao consumir, pois ao fornecedor é dada toda a responsabilidade do risco do negócio, portanto, incidentes de segurança afetam o consumo do serviço ou do produto e, por consequência, o consumidor.

Ressalta-se que em ambos os relacionamentos, tanto de consumo quanto de proteção de dados, são regidos pela boa-fé. Com isso, a informação suficientemente clara, objetiva e completa visa proteger sempre a parte de maior vulnerabilidade. Essa vulnerabilidade consiste em três aspectos quando se olha a relação de consumo/proteção de dados/Inteligência Artificial: a vulnerabilidade do consumidor, a vulnerabilidade do titular de dados e a vulnerabilidade tecnológica. Nos dois primeiros, a vulnerabilidade é presumida legalmente, a do consumidor expressamente no artigo 4º, inciso I do Código de Defesa do Consumidor e na proteção de dados implicitamente ao longo da lei, evidenciando-se o enfoque nos direitos dos titulares de dados, nos artigos 17 ao 22, dentre outros, na Lei Geral de Proteção de Dados.

A terceira atinge todas as relações que envolvem tecnologia, principalmente as que contêm implementação de Inteligência Artificial. A vulnerabilidade do usuário final é tangível, pois a tecnologia direciona e dirige o comportamento do usuário, limitando suas ações no ambiente digital e no analógico, fazendo-o consumir apenas o que lhe é apresentado ou acessível pela tecnologia. Quem desenvolve e opera tais tecnologias, por consequência, também direcionam e limitam o comportamento do usuário, pois aquele estabelece as regras de funcionamento de todo o espoco da tecnologia em sua operação. Por essa razão que a transparência é fundamental para lidar com este aspecto de vulnerabilidade.

A transparência, portanto, configura-se como o oposto da opacidade, fornecendo o conhecimento do que será e do que foi produzido em termos de atividades, processos e resultados. Já a opacidade se traveste da invisibilidade e ininteligibilidade, atrapalhando a compreensão dessas mesmas atividades, processos e resultados pelo titular de dados. A opacidade, por sua vez, traz problemas relevantes em alguns tipos de tratamento de dados, em especial os automatizados. Sistemas computacionais, por exemplo, são construídos para que seu usuário saiba apenas as entradas e saídas, ignorando a clareza no processamento intermediário.

Sistemas computacionais desenvolvidos com técnicas tradicionais de modelagem e programação, utilizando algoritmos determinísticos, possuem característica de previsibilidade de resultados, mesmo que não se saiba como aquele determinado resultado foi alcançado. Em modelos que se lidam com incerteza, como sistemas com Inteligência Artificial, pode não existir previsibilidade, pois se lida (em sua maioria) com a probabilidade e a estatística, onde nem sempre as mesmas entradas terão os mesmo resultados. Em ambos os casos, não há, por padrão, transparência no processo, caracterizando a denominada opacidade algorítmica, com exceção dos sistemas de fonte aberta, comumente denominados de *open source*.

A opacidade algorítmica, também conhecida como "caixa-preta" algorítmica, se refere à falta de transparência e compreensão sobre como os algoritmos são desenvolvidos, como funcionam e como são utilizados em determinados sistemas e aplicações (WACHTER; MITTELSTADT; RUSSELL, 2017). Isso gera dificuldades para que os usuários compreendam como seus dados pessoais são tratados em determinados processos automatizados, como aqueles realizados por sistemas de inteligência artificial.

A opacidade algorítmica pode ser um problema para a proteção de dados pessoais, pois dificulta a identificação de potenciais problemas e vieses que possam existir em determinado sistema ou processo automatizado. Por exemplo, um sistema de reconhecimento facial que utilize algoritmos que discriminem determinados grupos de pessoas, como negros ou mulheres, pode causar danos à privacidade e aos direitos humanos desses grupos. Destarte, se o algoritmo for opaco e o processo de desenvolvimento e treinamento do sistema não for transparente, pode ser difícil detectar e corrigir esses vieses que resultam em discriminação negativa.

Diante desse cenário, a preocupação de como esses tipos de ferramentas computacionais são produzidas chamou atenção do legislador ordinário, que propôs o Projeto de Lei 21/2020 e o 2338/2023, discutidos na Seção 3.2. Como visto na seção citada, as propostas trazem, em seu escopo, uma série de fundamentos, princípios (definidos no artigo 5º) e diretrizes para direcionar a utilização responsável das tecnologias que implementam técnicas de Inteligência Artificial. Nesses projetos, a transparência, contida no artigo 5º, inciso V no Projeto de Lei 21 de 2020, e no artigo 3º, inciso VI no Projeto 2338 de 2023, encontra-se como um princípio e este perpassa por todo o texto, influenciando diretamente na criação e uso de sistemas inteligentes. No texto dos projetos, essa transparência é definida como o direito de informação clara e precisa sobre o uso de tecnologias que implementam soluções de Inteligência Artificial, além de sua associação com a explicabilidade dos sistemas utilizados, contanto com exceção prevista em lei ou para

salvaguardar segredos comerciais e industriais.

Há a previsão, outrossim, de hipóteses em que a transparência na forma de dever de informar deve ser observada. Essas mesmas hipóteses residem nas alíneas a, b e c do mesmo inciso no qual a transparência está estabelecida, são elas: a) a informação sobre a comunicação direta com um sistema de Inteligência Artificial; b) a identificação do responsável pela operação do sistema inteligente, seja pessoa natural ou jurídica; e c) os critérios gerais de funcionamento do sistema de Inteligência Artificial, neste ponto devem ser observados também os segredos comerciais e industriais, sobretudo quando houver risco à direitos e garantias fundamentais.

Ainda sobre o texto do projeto de lei, existe uma harmonia com a proteção de dados, expressa no artigo 4ª, nos incisos VIII e XV especificamente. O artigo em questão aponta os fundamentos para o desenvolvimento e para a aplicação da Inteligência Artificial. No inciso VIII, a segurança, a privacidade e a proteção de dados pessoas são expressamente indicadas, enquanto no inciso XV se estabelece a harmonização com a lei de nº 13.709 de 2018, a Lei Geral de Proteção de Dados. Isso porque sistemas que possui Inteligência Artificial do qual se faz uso de aprendizado de máquina e aprendizado profundo se utilizam de volume considerável de dados e, por consequência, também fazem uso de dados pessoais tanto para treinamento quanto para testes (em ambiente de desenvolvimento) como também para a retroalimentação e extrapolação do que fora aprendido (em ambiente de produção).

Como ambos os textos se preocupam com a utilização desenfreada desse tipo de tecnologia, os dois prezam pela transparência como resposta à opacidade que o funcionamento da tecnologia em questão geralmente apresenta. Os textos também mostram a preocupação com segredos comerciais e industriais, revelando uma restrição quanto à transparência e ao que pode ser divulgado para terceiros, sendo que o molde mais apropriado em relação aos níveis de transparência apresentados nesta seção dependerá do caso concreto e do equilíbrio entre o direito à privacidade e a proteção de dados com o interesse legítimo das instituições em proteger informações sigilosas.

Diante do uso de sistemas com Inteligência Artificial nas atividades de negócio, a transparência total poderia prejudicar o desenvolvimento econômico e para as instituições privadas, sobretudo, a competitividade e saúde financeira. Isso tem particular relevância em modelos de negócios baseados no ambiente digital, onde a utilização desses sistemas forma a espinha dorsal da atividade. Por outro lado, a transparência limitada prejudicaria o titular de dados que, por mais que se objetive o cumprimento de obrigações legais e regulatórias, esse nível de transparência evidencia o estado de hipossuficiência das pessoas, deixando-as ainda mais vulneráveis ao tratamento automatizado, com incerteza de resultados e com poucas informações relevantes sobre o tratamento.

A transparência seletiva parece pender para a maior proteção dos agentes de tratamento do que para os titulares de dados. Entretanto, como poderá existir uma seleção de informações relevantes quanto ao sistema automatizado com intuito de direcionar um conhecimento resguardados os legítimos interesses dos agentes de tratamento, neste aspecto em específico,

é aceitável, desde que se cumpra as demais exigências do dever de informar sobre o tratamento, pois se evidenciam informações não técnicas sobre os sistemas. O cuidado com esse nível reside no fato dele servir como suporte de narrativa dos agentes de tratamento, direcionando a uma compreensão insuficiente do tratamento como um todo, assim como na transparência limitada.

Já na transparência parcial, há um equilíbrio de relações entre os agentes de tratamento e o titulares de dados. Esse equilíbrio ocorre porque a exceção é a retirada ou ocultação de informações por interesse legítimo, sendo, portanto, o padrão a divulgação de todas as demais informações relevantes. Outro fator determinante é o acesso restrito a essas informações. Enquanto na transparência seletiva, o acesso é permitido somente àqueles que possuem autoridade ou poder para tal, na transparência parcial, as informações são acessadas apenas por aqueles que possuem a devida autorização. Destarte, a transparência parcial, no entanto, parece ser o nível mínimo de transparência aceitável ao qual o titular de dados tem direito, por conta da possível restrição que resguarda os segredos comerciais e industriais apontados na lei.

Entretanto, no âmbito do Poder Judiciário, ao se utilizar ferramentas de software na representação processual, especialmente quando elas implementam técnicas de Inteligência Artificial, os níveis de transparência que apresentam alguma restrição não parecem viáveis em sua utilização, pois, deve existir a transparência total em como esses sistemas foram criados e como eles funcionam, uma vez que devem observar o devido processo legal. Neste sentido, a Resolução 332/2020 do Conselho Nacional de Justiça que regulamenta o uso de sistemas com Inteligência Artificial no Judiciário traz a transparência como requisito fundamental para o desenvolvimento e o uso destes tipos de sistemas.

No cenário processual, a transparência das ferramentas de software permite que o sistema possa ser auditado para que se garanta o funcionamento sem vieses que prejudicam as partes de alguma forma. Que se possa, igualmente, combater decisões criadas com auxílio de sistemas de Inteligência Artificial, sabendo-se como uma determinada decisão foi criada. Essa discussão será objeto do Capítulo 5, onde terá um aprofundamento no uso dessa tecnologia no Poder Judiciário.

Para Kaminski (2018), na comunidade acadêmica, tem havido uma demanda por transparência na tomada de decisões automatizadas. Essa transparência pode ser realizada através de notificações para indivíduos e auditorias que permitem supervisão de terceiros especializados. Parte da literatura defende uma transparência ampla e profunda, sugerindo que tanto o código-fonte algorítmico quanto os conjuntos de dados devem ser submetidos à análise pública. No entanto, outros argumentam que essa transparência pode causar danos ou que notificações direcionadas aos indivíduos podem ser relativamente inúteis, uma vez que eles não têm conhecimento técnico para interpretá-las de forma significativa.

Por fim, levando-se em consideração o ecossistema de proteção e privacidade de dados, a transparência deve tomar a forma de uma explicação significativa, orientando-a em sua construção e concretizando-a com intuito de substanciar de conhecimento relevante o destinatário dessa explicação, isto é, o titular de dados pessoais. Por esta razão, foi preciso

analisar os níveis de transparência e identificar os que poderiam se adequar ao contexto desse ecossistema. Portanto, passa-se a analisar o direito à explicação de decisões automatizadas na Seção 4.2.

4.2 A explicação significativa de decisões automatizadas como fundamento para o exercício de direitos dos titulares de dados pessoais

A explicação, em um cenário mais abrangente, consiste em uma descrição ou justificativa de algo que aconteceu ou da forma que funciona. Seria uma espécie de descrição da causa ou do motivo de uma situação ou fenômeno. Essa dita explicação pode ser contextualizada em diferentes maneiras, incluindo-se a científica, a filosófica, a sociológica, a tecnológica, a jurídica, entre diversas outras.

A explicação usualmente se baseia em conhecimento pré-estabelecido e em evidências, visando o esclarecimento de conceitos, processos e resultados. Além disso, a explicação pode ser utilizada para descrever e justificar uma determinada decisão ou ação, ou pode servir para amparar uma resposta a uma pergunta ou crítica. Em suma, a explicação consiste em uma ferramenta importante para a compreensão e resolução de questões, assim como para uma comunicação clara e eficaz.

Quando essa explicação é relacionada à proteção de dados pessoais, ela vem em forma de direito do titular de dados, da qual a explicação deve ser direcionada com transparência e significado. Isso porque a transparência e a informação significativa formam os pilares para uma explicação efetiva que vai além da mera descrição de determinados procedimentos.

Assim, a transparência pode ser utilizada para explicar o funcionamento de uma arquitetura de sistema ou modelo de representação de dados, a fim de descrever como um determinado sistema opera em casos específicos. Isso ocorre porque a informação significativa para a transparência não precisa abordar detalhes técnicos extremamente complexos, mas sim fornecer uma explicação adequada ao nível de compreensão médio de um ser humano não especializado na área.

Para que o titular de dados possa exercer seus direitos conforme estabelecido pela legislação, é essencial que as explicações fornecidas contenham informações significativas que permitam ao indivíduo entender como seus dados foram tratados e tomar decisões informadas em relação a eles. Além disso, as informações devem estar em conformidade com a legislação brasileira aplicável ao caso específico, permitindo que o titular possa buscar a correção de tratamentos que afetem seus direitos de forma positiva ou negativa.

Neste sentido, Selbst e Powles (2017) aponta a existência de dois caminhos para entender o valor de uma explicação dentro desse contexto: como instrumento ou de forma intrínseca (um aspecto fundamental da autonomia e da personalidade). Para os pesquisadores, ambos os caminhos são importantes, porém o instrumental oferece um molde concreto para medir se a explicação é significativa o suficiente.

Esse tipo de explicação, a priori, como instrumento, se aplica em qualquer tratamento que envolva dados pessoais, mas o real impacto incide sobre os tratamentos automatizados destes dados, o que abarca o tratamento realizado por sistemas com Inteligência Artificial. Esse último traz particular desafio em termos de explicação, uma vez que esses sistemas possuem um nível de opacidade elevado diante de sua arquitetura, o que tem como consequência a dificuldade em informar a causa de uma decisão específica em um tratamento de dados inteligente.

Isso se mostra importante porque o momento em que a explicação é realizada define que tipo de informações que devem ser fornecidas. Esses momentos podem ser categorizados como: 1) explicação ou avaliação antes de um determinado evento ou ação ter ocorrido, denominada *ex ante*; e 2) explicação ou avaliação depois de que um determinado evento ou ação ocorrida, *ex post* (KIM; ROUTLEDGE, 2018). Na primeira, há uma previsão ou estimativa com base em diversas fontes de informação, a exemplo dos modelos computacionais dos sistemas com Inteligência Artificial. Usualmente se espera obter conhecimento sobre as funcionalidades do sistema, ou seja, a lógica, significado, consequências previstas e funcionalidade geral de um sistema automatizado de tomada de decisão, por exemplo, a especificação de requisitos do sistema, árvores de decisão, modelos pré-definidos, critérios e estruturas de classificação.

A segunda, há a análise retrospectiva que procura avaliar o desempenho ou resultado de uma ação ou evento já ocorrido. O principal objetivo da explicação *ex post* é compreender as causas e efeitos do evento ou ação, identificar pontos fortes e fracos, e aprender com o passado para melhorar o futuro, além da lógica envolvida na decisão específica, razões e circunstâncias individuais de uma decisão automatizada específica, por exemplo, a ponderação de recursos, regras de decisão específicas de caso definidas por máquina, informações sobre referências ou grupos de perfis. Observa-se que ambas as abordagens são complementares, enquanto uma explica toda a estrutura e regras que podem ser utilizadas antes de um evento como uma decisão, a outra explica as causas que levaram a uma determinada decisão. Entretanto, nada impede que uma explicação *ex post* venha acompanhada por uma explicação *ex ante* primeiramente.

Dessa forma, segue o Tabela 2 com o comparativo das características das explicações *ex ante* e *ex post* através do momento da explicação, finalidade, foco, tipo de informação, precisão, nível de detalhamento e uso.

Tabela 2 – Comparativo de explicação *ex ante* e *ex post.*

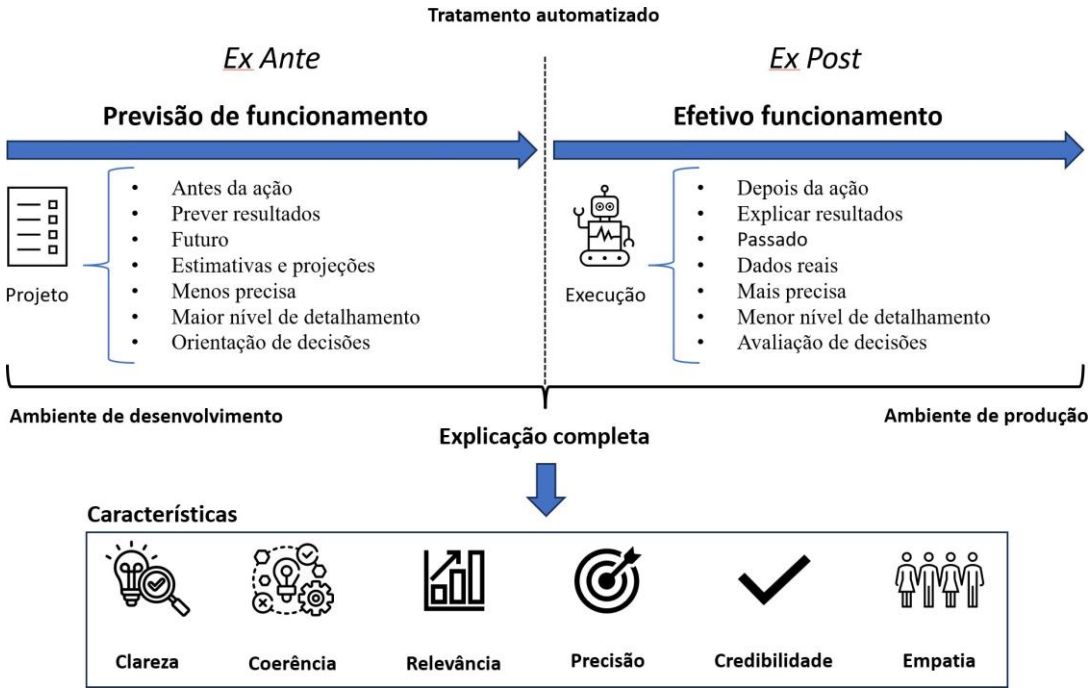
Características	***Ex Ante***	***Ex Post***
Momento da explicação	Antes da ação	Depois da ação
Finalidade	Prever resultados	Explicar resultados
Foco	Futuro	Passado
Tipo de informação	Estimativas e projeções	Dados reais
Precisão	Menos precisa	Mais precisa
Nível de detalhamento	Maior nível de detalhamento	Menor nível de detalhamento

Uso	Orientação de decisões	Avaliação de decisões

Fonte: adaptada de Selbst e Powles (2017), Kim e Routledge (2018).

No Tabela 2, é possível visualizar de forma sintética o propósito desses tipos de explicação. O principal ponto de diferença entre eles reside no fato de existirem em momentos diferentes. A Figura 13 mostra a relação entre as explicações *Ex Ante* e *Ex Post* com as características de uma explicação eficiente.

Figura 13 – Explicação Completa.



Fonte: O autor.

Embora existam momentos distintos sobre uma explicação, ambas precisam de características que forneçam eficiência e efetividade (SELBST; POWLES, 2017). Alguns destes atributos são:

- **Clareza**: uma explicação eficiente deve ser clara, concisa e de fácil entendimento. A pessoa ou entidade no papel de emissor da explicação deve levar em consideração quem receberá a mensagem, pois se assim não o fizer o risco de se ter uma explicação inócua é significativo.
- **Coerência**: uma explicação coerente tem como padrão uma sequência logicamente encadeada de ideias, o que permite que o receptor compreenda a mensagem de forma fácil e organizada.

- **Relevância**: uma explicação deve ser relevante para o receptor, trazendo em seu escopo os elementos necessários e de seu interesse. A relevância deve observar o tipo de linguagem empregada na mensagem, assim como considerar elementos claros e coerentes.
- **Precisão**: uma explicação precisa ser exata e evitar imprecisões e ambiguidades. É importante que o emissor escolha as palavras certas para transmitir a mensagem de maneira adequada.
- **Credibilidade**: uma explicação precisa ser confiável e baseada em fontes confiáveis. O emissor deve ter autoridade e conhecimento sobre o assunto que está explicando, para que o receptor confie na mensagem.
- **Empatia**: uma boa explicação deve ser empática, ou seja, o emissor deve considerar as necessidades do receptor e adaptar a explicação a essas necessidades. É importante que o emissor se coloque no lugar do receptor e transmita a mensagem de forma que ele possa entender.

Portanto, quando uma explicação é clara, coerente, relevante, precisa, credível e empática, ela é mais facilmente compreendida e assimilada pelo receptor. Com base na explicação fornecida com tais características, o receptor poderá tomar a ação mais adequada ou não agir se assim não lhe for conveniente. A falta de algum desses atributos atrapalha o receptor na sua compreensão ao passo que pode atrapalhar em sua tomada de decisão.

De mesma forma, deve-se ter cuidado com o tipo de conteúdo que será utilizado para explicar algo, pois informações sensíveis para o negócio e segredos industriais precisam ser protegidos e sua exposição pode prejudicar a atividade de uma entidade. Além disso, explicações excessivamente técnicas podem não ser compreendidas por um receptor não técnico, perdendo-se o objetivo primário de uma explicação significativa.

Nesta perspectiva, até o formato de apresentação da explicação pode fazer diferença. As técnicas de *Visual Law* podem facilitar essa apresentação, pois agregam elementos complexos, como cláusulas contratuais ou explicações complicadas em formato textual, com elementos gráficos, como imagens e fluxogramas com o fim de tornar mais acessível para quem não tem conhecimento jurídico. O *Visual Law* tem como objetivo tornar o direito mais acessível, reduzir a complexidade e melhorar a compreensão das questões legais, tornando-o mais inclusivo para pessoas que não têm formação jurídica. Isso pode ser particularmente importante em um mundo onde questões legais desempenham um papel cada vez mais significativo em muitos aspectos da vida cotidiana, como contratos, disputas e direitos individuais.

Apesar do *Visual Law* ser uma estratégia de apresentação de informações complexas de maneira mais simplificada, ela deve ser utilizada com outras técnicas em conjunto, pois as pessoas com deficiência visual usualmente se utilizam de sistemas que os ajudam a realizar a leitura de documentos textuais. Ler e converter em áudio o que se está escrito é muito mais factível de se implementar do que criar sistemas de visão computacional capazes de reconhecer todo o tipo de imagem, sem estrutura prévia nenhuma associada, em áudio que explique o que

se quer passar.

O impacto de uma explicação exclusivamente técnica sobre determinado tratamento automatizado por Inteligência Artificial se aproxima dos efeitos que a opacidade tem nestes sistemas. A opacidade, como dito na Seção 4.1, está em oposição à transparência e, quando aplicados em sistemas computacionais, ela se caracteriza como a dificuldade de entender o funcionamento de sua arquitetura (envolvendo algoritmos, representação de dados, elementos externos e internos de comunicação com outros sistemas, etc.) (FERRARI, 2018). Quando a opacidade está nos sistemas com Inteligência Artificial, devido a implementação de algoritmos complexos e a utilização de um volume de dados considerável, esses sistemas se tornam difíceis de compreender e, por consequente, de explicar.

As dificuldades advindas do prisma tecnológico se caracterizam como um peculiar desafio para a explicabilidade do tratamento realizado por Inteligência Artificial. A começar pelas técnicas modernas de implementação da tecnologia que traz o elemento de complexidade na explicação *ex Post* de um determinado tratamento. Isso porque na explicação *ex ante* é possível aplicar os níveis adequados de transparência esperados pela legislação, pois se conhece o modelo utilizado, a arquitetura montada e o processo de negócio vinculado à tecnologia em questão.

Uma explicação *ex ante* é perfeitamente possível de se realizar porque os projetistas sabem os requisitos de negócio e como traduziram isso para composição do software de computador, ou seja, os projetistas saberão informar como o sistema funcionará em nível conceitual. Dessa forma, poderão juntamente com os idealizadores do sistema, definir quais as informações poderão ser disponibilizadas para as partes interessadas, como os titulares de dados, a fim de resguardar os seus segredos de negócio.

A principal preocupação com a explicação *ex ante* reside na forma de comunicação com o receptor, que deve conter todas as características de uma explicação eficiente e efetiva. As informações prestadas que fazem referência ao sistema inteligente deverão ser simplificadas na medida que a complexidade do sistema for crescendo. Essa simplificação estende-se aos dados e a preparação deles para adentrarem no sistema. Usualmente, os dados são fruto de uma construção chamada de Big Data, um volume considerável de dados de fontes diversas e que possuem uma grande variabilidade, precisando de um pré-processamento que visa sanitizar os dados para compor a base de conhecimento em sistemas inteligentes.

A composição dos dados também precisará ser explicada por ser a base de funcionamento das técnicas mais modernas de Inteligência Artificial. Os dados precisam ser localizados, agrupados, extraídos, filtrados, transformados para um padrão, armazenados para serem analisados e, a partir disso, adentrarem em uma base de dados de treinamento ou de testes para que a camada de raciocínio passa usá-los.

Com os dados preparados e sanitizados, o foco se volta para o modelo que será utilizado. Existem inúmeros modelos de algoritmos de Inteligência Artificial que podem ser utilizados para os mais diversos fins, a depender da tarefa que será realizada. Esses modelos são testados

em conjunto com os dados para a escolha do modelo que oferece o resultado com precisão mais significativa. Explicar os modelos que compõe determinado sistema é perfeitamente viável. Isto porque para construí-los é necessário compreendê-los.

A Figura 8 consiste em um exemplo visual simplificada que explica uma arquitetura de um sistema com Inteligência Artificial que pode impactar o ser humano. Desta forma, um modelo complexo tem sua explicação reduzida para que pessoas não técnicas possam compreender o funcionamento geral de um sistema inteligente genérico, sem especificações que poderia ser complementada também de uma maneira mais legível para este público.

Ainda na explicação *ex ante*, explicar o nível de autonomia, assim como a abordagem de aprendizado de máquina adotada também não se configura como um problema relevante, pois, repetindo o que ocorre com o modelo, se faz necessário entender para construir, o que remete a fase de desenvolvimento do sistema, pois há um projeto que direciona todas as escolhas da equipe de desenvolvimento. Isso coaduna com o que afirma Doshi-Velez et al. (2017), que diz que existe uma grande variação em torno de quando a explicação é exigida, mas também existem consistências importantes: ao exigir explicação de humanos, o que normalmente se quer saber é se e como certos fatores de entrada afetaram a decisão final ou o resultado. Uma explicação *ex ante* fornece o que se espera, neste sentido, mas não de forma completa, exigindo seu complemento em termos *ex post*.

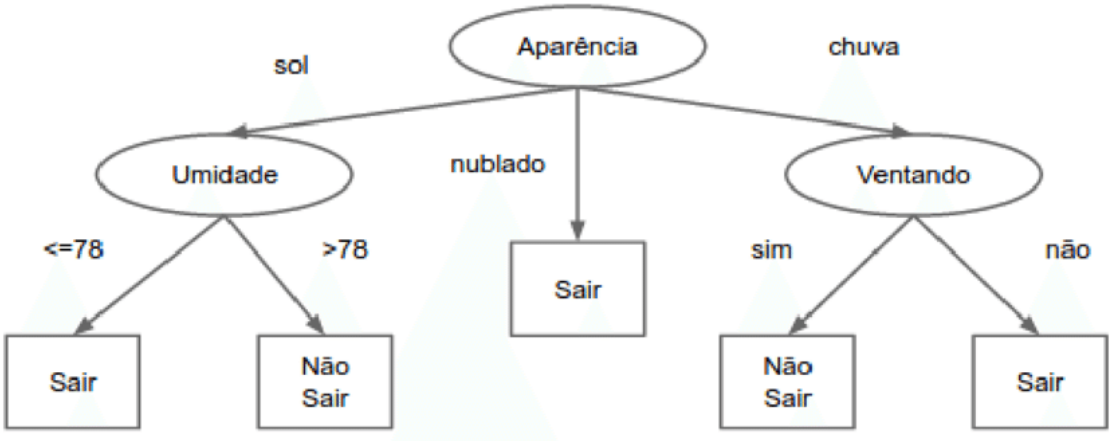
Ainda para Doshi-Velez et al. (2017), a possibilidade de fornecer explicações juridicamente aceitáveis é viável e se baseia no fato de que explicação e transparência são conceitos diferentes. No estudo dos autores, para fornecer uma explicação não é necessário compreender o fluxo de dados em um sistema de Inteligência Artificial, assim como para explicar o processo cognitivo humano não é necessário compreender o fluxo de sinais neurais, que não são interpretáveis para um ser humano. Em vez disso, a explicação, conforme exigido pela lei, trata de responder como certos fatores foram usados para chegar ao resultado em uma situação específica. A premissa dos autores, no entanto, só poderia ser testável em casos que já existam decisões concretas, pois os fatores identificados que influenciam o resultado seriam checados, não sendo suficientes a previsão conceitual desses.

Por outro lado, a explicação *ex post* requer uma análise cuidadosa dos eventos que já ocorreram dentro do sistema. Ela pretende explicar o processamento já realizado pelo sistema inteligente, trazendo em seu escopo, uma série de informações que dependem do entendimento técnico do fluxo de dados dentro de modelos complexos. Neste tipo de explicação, até a abordagem de aprendizado de máquina é relevante. Isso pois, na prática, a tomada de decisão automatizada geralmente envolve criação de perfil, onde os perfis orientam o processo de tomada de decisão (KAMARINOU; MILLARD; SINGH, 2016). Então há, na maior parte das operações que resultam em tomadas de decisão, subprocessos de classificação que definem perfis e a partir destes uma saída é direcionada. A Figura 8 igualmente pode ser usada para exemplificar esse procedimento de classificação, dentre outros subprocessos, antes de uma tomada de decisão.

Quando se foca na explicação *ex post* de um sistema complexo que utiliza técnicas

de aprendizagem de máquina, há gradações de dificuldade a depender do modelo utilizado. A título de exemplo, considera-se uma estrutura de árvore de decisão para a classificação de alguma situação. A explicação de como o sistema atinge determinada resposta com base em determinados dados de entrada não oferece complexidade de opacidade que impeça o entendimento do fluxo de dados da entrada até a saída do modelo. A Figura 14 mostra exemplo de um modelo simples baseado em uma árvore de decisão[1] para sair ou não de determinado lugar.

Figura 14 – Árvore de decisão para sair ou não de determinado local.



Fonte: Ribeiro (2015).

Na Figura 14 visualiza-se uma estrutura para a tomada de decisão que auxilia o tomador da decisão, seja uma máquina ou um ser humano, em sair de um determinado local ou não, com base na umidade, no vento e no clima (sol, chuva ou nublado). Esse modelo é simples de ser implementado por um sistema de Inteligência Artificial e sua saída (Sair ou Não Sair) pode ser auditado até a origem do processamento, ou seja, não existe dificuldade na explicação do processo de tomada de decisão.

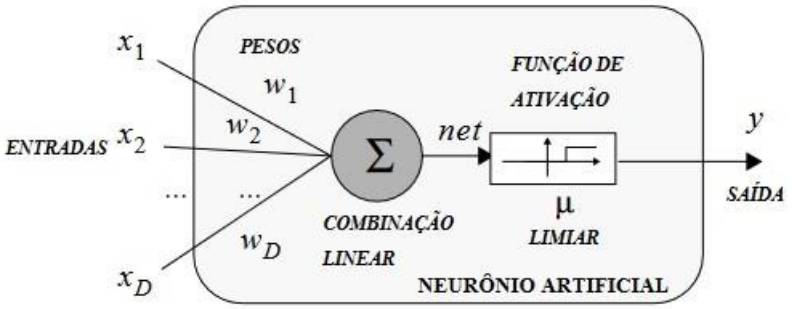
Em oposição, ao utilizar técnicas mais avançadas, como as redes neurais artificiais, a dificuldade mencionada no parágrafo anterior se torna mais evidente. Segundo Aggarwal (2018), as redes neurais artificiais são modelos computacionais que se inspiram no funcionamento do cérebro humano para realizar tarefas de aprendizado de máquina. Elas consistem em um conjunto de neurônios artificiais interconectados, organizados em camadas, que são capazes de aprender e generalizar a partir de um conjunto de exemplos. As redes neurais são amplamente utilizadas em reconhecimento de padrões, classificação de imagens, previsão de séries temporais, entre outras aplicações.

A Figura 15 apresenta a estrutura de um neurônio artificial em um formato conceitual baseado na matemática. O modelo ilustrado expõe apenas um único neurônio. Esse neurônio se

[1]Uma árvore de decisão é uma técnica de modelagem preditiva que utiliza uma estrutura em forma de árvore para representar um conjunto de decisões e suas possíveis consequências. É uma ferramenta de análise que auxilia na tomada de decisões em situações complexas e incertas, ao mapear todas as possíveis opções e seus resultados esperados (QUINLAN, 1986).

configura como o menor elemento de uma rede neural artificial.

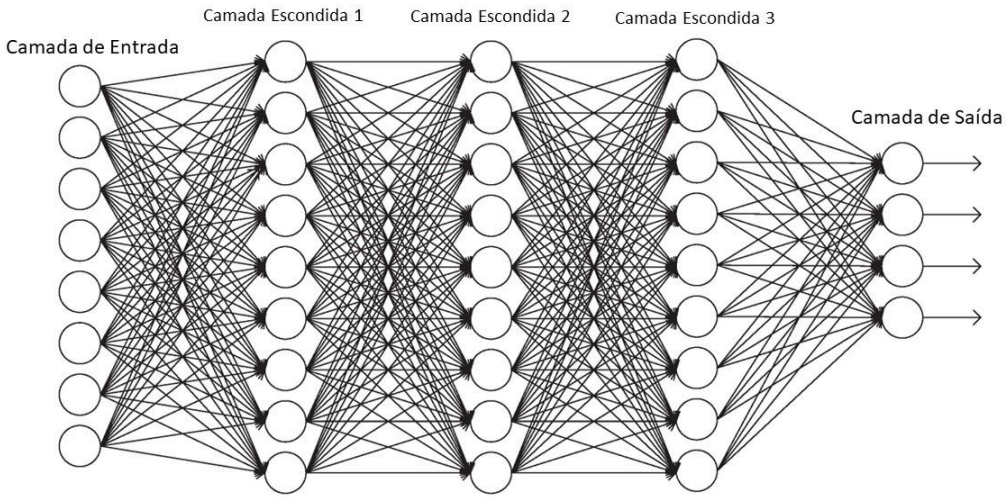
Figura 15 – Neurônio Artificial - modelo *Perceptron*.



Fonte: Rauber (2005).

Na Figura 15 é possível visualizar as entradas no modelo de simulação do neurônio, assim como seu processamento indicado pelos valores que representam pesos (valores que serão multiplicados com as entradas), um somatório dos valores de entradas já multiplicados por seus respectivos pesos, e a função de ativação do neurônio, que tem como entrada o resultado do somatório anterior. Por fim, se tem a saída do neurônio, resultado do cálculo da função de ativação. O modelo em si é de fácil compreensão, no entanto, quando agrupados com outros neurônios artificiais, seu entendimento em termos de processamento se torna difícil de se explicar. A Figura 16 mostra o modelo de uma rede neural artificial aplicável em diversos cenários.

Figura 16 – Rede Neural Artificial Profunda.



Fonte: Baseada em Cetax (2022).

A Figura 16 exemplifica um conjunto de neurônios artificiais - como o exposto na Figura 15 - em rede, contendo diversas camadas até a saída de seu processamento. Uma situação em que o modelo não pode ser explicado ocorre quando os caminhos que levaram à sua saída não são facilmente discerníveis durante um processamento específico. Importante ressaltar que os modelos de aprendizagem de máquina aqui mostrados nas Figuras 14, 15 e 16 possuem previsibilidade em sua saída, pois adotam a abordagem supervisionada (Capítulo 2, Subseção 2.3.2), isto é, as saídas dos modelos são conhecidas e pré-estabelecidas desde o projeto.

A dificuldade trazida pela opacidade no processamento pode se manifestar como uma barreira tecnológica para explicar a execução do modelo em um contexto real, em ambiente de produção de resultados. Essa opacidade se torna ainda mais problemática quando se implementa uma abordagem de aprendizagem de máquina não supervisionada. Isso porque não há previsibilidade em suas saídas até que exista o processamento do modelo, pois ele mesmo irá encontrar padrões e definir com base nos dados as saídas mais adequadas. O problema em explicar tais situações se agrava ainda mais com esse tipo de abordagem de aprendizagem de máquina.

Segundo Kim (2017), os anseios dos titulares de dados perante uma explicação vinda de um tratamento como este deve atender alguns questionamentos que os inquietam, são eles: qual foi a contribuição dos principais fatores em uma decisão? Essa é a interpretação mais frequente de uma explicação para uma decisão. Em diversas situações, a sociedade estipula uma lista de fatores que devem ser ou não levados em conta em uma decisão específica; alterar um fator específico teria mudado a decisão? Às vezes, o que se deseja saber não é se um fator foi considerado, mas se ele foi determinante. Isso é especialmente útil quando o tomador de decisão tem acesso a uma informação que tem usos adequados e inadequados, como o uso da raça nas admissões universitárias. Ao analisar o efeito da alteração dessa informação no resultado e compará-lo com as expectativas, é possível inferir se foi utilizada corretamente; e por que dois casos semelhantes resultaram em decisões diferentes ou vice-versa? Por fim, é possível querer saber se um fator específico foi determinante em relação a outra decisão. Essa informação é útil quando é necessário avaliar a consistência e a integridade do tomador de decisão.

A explicação que abarca esses elementos, seja ex ante ou ex post, tem por objetivo esclarecer todo o processo de tratamento de dados pessoais realizados por sistemas dotados de Inteligência Artificial. Desafios de cunho tecnológico são perceptíveis quando analisa-se as principais técnicas e modelos utilizados para a aprendizagem de máquina e extrapolação de conhecimento representado em um sistema. Explicar o processamento específico pode ser uma tarefa difícil ou até mesmo impraticável. O como deve ser a explicação, diante de todo o exposto, para se amoldar o que é definido na lei de forma efetiva é o foco da próxima seção.

4.3 O direito à explicação como garantia para o titular de dados frente ao tratamento automatizado por Inteligência Artificial

A vulnerabilidade do titular de dados é um aspecto crucial que se torna evidente com o uso crescente de sistemas baseados em Inteligência Artificial. Essa vulnerabilidade se manifesta em vários níveis, principalmente devido à assimetria de poder e conhecimento entre os indivíduos que detêm o controle sobre o desenvolvimento e operação desses sistemas e aqueles que são afetados por suas decisões.

Essa assimetria de poder tecnológico pode resultar em violações de direitos e garantias fundamentais dos titulares de dados. Por exemplo, sistemas de Inteligência Artificial são usados em processos de seleção de emprego podem discriminar candidatos com base em critérios não transparentes, como gênero, raça ou idade, sem que os candidatos tenham meios eficazes para contestar essas decisões injustas. Isso viola princípios de igualdade, equidade e não discriminação, dentre outros direitos fundamentais basilares em um Estado democrático de direito.

O direito à explicação de decisões automatizadas é crucial em um cenário onde a inteligência artificial desempenha um papel crescente em decisões que afetam profundamente a vida das pessoas. Primeiramente, ele promove a transparência e a prestação de contas, permitindo que os indivíduos compreendam como uma decisão foi tomada e, quando necessário, responsabilizem as partes envolvidas por erros ou discriminação. Além disso, esse direito é fundamental para prevenir a discriminação e os vieses em decisões automatizadas, possibilitando que os indivíduos identifiquem qualquer discriminação e contestem essas decisões.

O direito à explicação também protege os direitos e liberdades fundamentais, pois muitas decisões automatizadas têm impacto direto nesses direitos, como acesso a crédito, emprego e liberdade pessoal. Ele desempenha um papel importante na melhoria contínua dos sistemas de Inteligência Artificial, permitindo que os usuários forneçam *feedback* para aprimorar a precisão dos algoritmos. Além disso, a transparência nas decisões automatizadas constrói a confiança do usuário, aumentando a aceitação e o uso desses sistemas. No âmbito jurídico, o direito à explicação é essencial para garantir que as partes afetadas entendam as razões por trás de uma decisão automatizada, permitindo uma defesa adequada. Por fim, ele reflete princípios éticos e responsáveis no desenvolvimento e uso da inteligência artificial, comprometendo-se com a proteção dos direitos individuais em um mundo cada vez mais impulsionado por essa tecnologia.

De início, essa automatização do tratamento de dados pessoais ocorre através das tecnologias digitais, das quais atualmente são o suporte do funcionamento da Inteligência Artificial, como discutido no Capítulo 2. Não há como ignorar a necessidade de explicar também o aspecto tecnológico desse tipo de tratamento. As explicações tecnológicas nos aspectos *ex ante* e *ex post* devem esclarecer o funcionamento do software, mas precisam igualmente do complemento quanto aos requisitos do porquê o tratamento é realizado e quais são as justificativas legais para realização deste.

A transparência e o dever de informação estão intrinsecamente conectados à explicação, como descritos nas Seções 4.1 e 4.2. Dito isso, espera-se que esses elementos estejam contidos, mesmo que implicitamente, na concretização de um direito à explicação de um tratamento automatizado. Essa explicação significativa deste tipo de tratamento requer a observação do nível de transparência adequado ao que se pede no instrumento normativo - a Lei Geral de Proteção de Dados Pessoais, no já mencionado, artigo 20, §1º.

Em uma primeira análise, o direito à explicação de decisões automatizadas está estritamente ligado ao direito à revisão de tratamento automatizado regrado pelo artigo 20, do qual define que é garantido ao indivíduo o direito de pedir uma nova análise em relação às decisões que foram tomadas exclusivamente com base no uso de ferramentas automatizadas, as quais possam ter afetado seus interesses pessoais, profissionais, de consumo ou crédito, assim como os aspectos que compõem sua personalidade, incluindo a elaboração de seu perfil. Assim sendo, se pode intuir que o direito à explicação de decisões automatizadas só teria substância para apoiar esse direito à revisão.

Todavia, o direito à explicação pode ser exercido de forma autônoma, com o objetivo de suprir a necessidade do titular de dados em conhecer e entender como foi ou será realizado o tratamento automatizado. O titular de dados não detém obrigação alguma em se utilizar do direito à revisão de decisões automatizadas caso não deseje exercer. O mero desejo em compreender como seus dados serão processados por ferramenta automatizada é suficiente para exercer o direito à explicação.

Neste sentido, Barbosa, Toniazzo e Ruaro (2021), afirmam que é inegável que o artigo 20 da Lei Geral de Proteção de Dados consolidou a importância da transparência e da relação com o princípio da motivação decisória algorítmica. Essa consolidação pode ser vista como uma extensão da necessidade de fundamentação para o contexto das decisões artificiais.

Isso por conta da autodeterminação informativa, estabelecida no artigo 2º, inciso II da Lei Geral de Proteção de Dados, que se refere ao direito das pessoas de terem controle sobre seus próprios dados pessoais e de decidirem como e para quais finalidades esses dados serão utilizados. Em outras palavras, trata-se do direito de cada indivíduo de determinar quais dados pessoais ele deseja fornecer, quem pode ter acesso a eles, com que finalidade serão utilizados e por quanto tempo serão mantidas. Esse conceito é fundamental para a proteção da privacidade e da intimidade dos indivíduos em um mundo cada vez mais digital e conectado. A autodeterminação informativa permite que as pessoas tenham o controle sobre suas informações pessoais, o que é essencial para garantir sua liberdade e dignidade.

Por essa razão, se encontra no texto de lei que o controlador deve proporcionar "informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada" (BRASIL, 2018). O texto em si não aponta quais informações devem ser fornecidas e o quão aprofundadas elas devem ser, desde que sejam claras e adequadas e mesmo estes critérios não estão especificados.

Essa falta de especificação pode causar efeitos negativos, pois deixa nas mãos do

controlador estabelecer quais critérios ele acredita ser o suficiente e o quão claras e adequadas estão as informações que serão repassadas ao titular de dados. O fato de se ter a expressão "observados os segredos comercial e industrial" ao final do §1º do artigo 20, da lei em discussão, leva a crer que há uma limitação na explicação que pode funcionar como um escudo para o controlador, do qual ele pode recorrer sempre que não desejar ou não conhecer a fundo do processo de tratamento automatizado, esquivando-se, portanto, de uma obrigação legal.

Vale destacar que em caso o controlador recorra aos argumentos de segredo comercial e industrial para não fornecer informações necessárias ao titular de dados para que este possa compreender o tratamento, a Autoridade Nacional de Proteção de Dados pode realizar auditoria no controlador com o propósito de garantir que não tenha ocorrido nenhum tipo de discriminação negativa no tratamento. Isso é possível por conta do §2º, do artigo 20, da Lei Geral de Proteção de Dados. Porém, até o momento da conclusão deste trabalho, não existe regulamentação visando esse tipo de procedimento junto à citada autoridade.

A clareza da informação se refere à transparência, enquanto a adequação está na significação da informação. A primeira diz respeito a capacidade de tornar as informações acessíveis, compreensíveis e claras o suficiente para que o titular possa entender como a decisão automatizada foi tomada. Já a segunda significa que as informações devem ser relevantes e suficientes para que o titular possa entender como a decisão foi tomada e avaliar sua precisão e confiabilidade. Além disso, as informações devem ser apresentadas de forma apropriada e adaptada ao conhecimento e à compreensão do titular.

A combinação de ambos os aspectos é essencial para fornecer as informações claras e adequadas sobre os critérios e procedimentos utilizados para a decisão automatizada, como preconiza o texto de lei. Sem transparência e informações significativas, a explicação fornecida ao titular pode ser insuficiente para garantir seu direito à autodeterminação informativa e ao controle sobre seus próprios dados pessoais, como já discutido nas subseções anteriores deste capítulo.

Essa discussão é importante porque parte da hipótese de trabalho tem como base a explicação de decisões automatizadas por Inteligência Artificial. Neste sentido, a explicação do tratamento automatizado perpassa pelo problema da opacidade dos sistemas inteligentes e pela limitação apontada para proteção da atividade do controlador. Em relação à transparência, remete-se ao já analisado neste capítulo e dentre os níveis elencados na Seção 4.1, o nível de transparência parcial mostra-se adequado por apresentar o equilíbrio entre a proteção já mencionada para o negócio e a devida exposição das informações que refletem o tratamento de dados em foco.

Essa transparência atrai informações significativas específicas do tratamento, além do funcionamento dos sistemas, tais como: as informações de identificação do processo de negócio, os tipo de dados pessoais de entrada do processo, a finalidade do processamento, a base legal para o processamento, as informações de existência do relatório de impacto de proteção de dados referentes ao processo, as fonte dos dados, o local de armazenamento dos

dados, o compartilhamento dos dados, a transferência internacional, as medidas de segurança aplicadas, o tempo de retenção e os direitos dos titulares associados ao tratamento em específico (EDWARDS; VEALE, 2018).

Todos os tipos de dados levantadas no parágrafo anterior pode ser associado a uma explicação significativa ex ante complementada com o descrição da arquitetura do sistema, o que inclui o modelo contendo os algoritmos utilizados e a representação de conhecimento composto pelos dados também em uso. Uma explicação que incluem somente as entradas e saídas (quando predefinidas) adentram na explicação ex ante, como também analisado na Seção 4.2, assim como as consequências de cada resultado de saída do sistema.

As informações contidas na explicação significativa, defendido por Selbst e Powles (2017), tem como objetivo da explicação se estende a fornecer informações acionáveis sobre a tomada de decisões, de modo que as partes afetadas possam aprender se e como podem alcançar um resultado diferente. Explicações são valiosas, nessa visão, porque capacitam as pessoas a navegar efetivamente pelo processo de tomada de decisão. Tais crenças são evidentes nos requisitos de notificação de ações adversas das regulamentações de pontuação de crédito, mas passaram a dominar os debates mais recentes sobre a função regulatória de exigir explicações das decisões orientadas por modelos de forma mais geral.

Já a explicação *ex post* é um desafio na questão da transparência de processamentos automatizados de dados. É difícil descrever como determinado processamento foi realizado devido à arquitetura de software e modelos utilizados, como exemplificado na Figura 16. A solução definitiva ainda não existe, uma vez que depende da evolução da tecnologia para tornar possível a explicação do funcionamento dos modelos que atualmente são opacos.

A literatura tem abordado esse problema de opacidade com certa tendência, como mostra a Figura 26 do Apêndice A, e embora não se possa afirmar a existência de uma correlação definida entre esse problema e as abordagens de pesquisa por áreas, a Figura 25, também do Apêndice A, indica que o aspecto tecnológico tem sido trazido em conjunto com pesquisas jurídicas, sugerindo essa dependência de evolução da tecnologia com indicação de resoluções provisórias com a finalidade de contornar essa limitação. Em resumo, a transparência e a adequação das informações são fundamentais para garantir o direito à explicação dos processamentos automatizados de dados, e a evolução da tecnologia é essencial para aprimorar a capacidade de explicação *ex post* desses processamentos.

Em virtude da complexidade envolvida na tecnologia, bem como a necessidade constante de avanços na área, a explicabilidade em tempo de execução de sistemas automatizados de tomada de decisão torna-se um desafio. Nesse contexto, uma possível solução seria a submissão da revisão de uma tomada de decisão automatizada por seres humanos. Essa revisão humana permitiria uma explicação mais clara e adequada dos critérios e procedimentos utilizados na decisão automatizada, cumprindo o que é requerido pela lei. Além disso, essa abordagem também poderia mitigar possíveis erros ou vieses nos algoritmos e nos sistemas utilizados para a tomada de decisão, garantindo uma maior confiabilidade nas decisões.

No entanto, é importante ressaltar que essa solução não é isenta de desafios, como a necessidade de expertise por parte dos revisores humanos e a possibilidade de que suas próprias subjetividades possam influenciar na revisão, além da inviabilidade de proceder com uma revisão humana a depender do modelo de negócio do controlador pelo volume de decisões tomadas.

Anteriormente previsto na Lei Geral de Proteção de Dados, mais precisamente no §3º do artigo 20, havia a possibilidade de revisão humana em casos de decisões automatizadas. Contudo, a Lei nº 13.853 de 2019 excluiu essa obrigatoriedade legal. Isso significa que agora a decisão de realizar a revisão por seres humanos fica a critério do controlador de dados, que poderá optar por esse caminho caso seja viável, ou ainda submeter a nova análise dos dados do titular a um sistema ou outro método automatizado. Essa mudança na legislação traz consigo uma maior responsabilidade para os controladores de dados, que devem avaliar a necessidade e viabilidade da revisão humana, bem como garantir que o processo de tomada de decisão seja compatível com a lei.

Deve-se evidenciar que o Projeto de Lei 2338 de 2023, com relação à explicação, estabelece igualmente um direito à explicação sobre a decisão, recomendação ou previsão tomada por sistemas de inteligência artificial, no artigo 5º, inciso II. A pessoa afetada por sistema de inteligência artificial, como descrito no artigo 7º e 8º do projeto, poderá solicitar explicação sobre a decisão, previsão ou recomendação, com informações a respeito dos critérios e dos procedimentos utilizados, assim como sobre os principais fatores que afetam tal previsão ou decisão específica, incluindo informações sobre a racionalidade e a lógica do sistema, o significado e as consequências previstas de tal decisão para a pessoa afetada, o grau e o nível de contribuição do sistema de inteligência artificial para a tomada de decisões, os dados processados e a sua fonte, os critérios para a tomada de decisão e, quando apropriado, a sua ponderação, aplicados à situação da pessoa afetada, os mecanismos por meio dos quais a pessoa pode contestar a decisão e a possibilidade de solicitar intervenção humana, nos termos desta Lei.

O projeto de lei mencionado no parágrafo anterior, permite expressamente em seu artigo 8º, inciso V, portanto, que exista intervenção humana em uma possível revisão de um resultado advindo de um sistema com Inteligência Artificial, ao contrário da Lei Geral de Proteção de Dados, onde não há essa previsão, portanto, caso o Projeto de Lei 2338 de 2023 se converta em sei, os titulares de dados poderão se valer do regramento sobre Inteligência Artificial para substanciar seu pedido de revisão humana e não depender da boa vontade do controlador ou operador para tanto.

Pontos relevantes residem nos artigos 7º e 8º, inciso I do projeto em discussão, pois eles preveem especificamente requisitos operacionais para a concretização efetiva da explicação do tratamento automatizado por Inteligência Artificial. Novamente se atrai uma explicação *ex ante* e *ex post*, da qual já fora objeto de análise na Seção 4.2, no que tange ao aspecto técnico computacional da atividade automatizada, tanto em fornecer informações prévias (ênfase no artigo 7º) quanto para informações posteriores à resposta do sistema (ênfase no artigo 8º). Além

disso, o projeto traz o instrumento da avaliação de impacto algorítmico do qual pode fornecer uma visão geral do sistema em seu funcionamento, fundamentos e justificativas de seu uso, como também os riscos e mecanismos para tratá-los.

No entanto, não bastará definir um direito à explicação de decisões automatizadas, será necessário igualmente fiscalizar o seu cumprimento. Neste ponto tem-se a Autoridade Nacional de Proteção de Dados como uma autarquia de natureza especial, vinculada ao Ministério da Justiça e da Segurança Pública, criada para zelar pela proteção de dados pessoais no Brasil e responsável por regulamentar temas relacionados à proteção de dados pessoais, da qual compete, nos termos do artigo 55-J, inciso IV, a fiscalização e a aplicação de sanções em caso de tratamento de dados em desconformidade com a legislação, além de estabelecer orientações e pareceres sobre tratamento de dados pessoais no território brasileiro.

A autoridade em questão tem competência para intervir quando o tratamento automatizado, seja por sistemas tradicionais ou com implementação de Inteligência Artificial, consumir dados pessoais. Essa relação possui uma proximidade de tal significância que a autoridade elaborou um documento de análise preliminar sobre o projeto de lei 2338 de 2023. Segundo a autoridade, O Projeto de Lei de nº 2338 de 2023 possui diversos pontos de interação com a Lei Geral de Proteção de Dados. O texto destaca três campos de correspondência entre o projeto de lei e a Lei de proteção de dados que devem ser destacados, dada a possibilidade de eventuais convergências, sobreposições e conflitos com as atribuições legais da Autoridade Nacional de Proteção de Dados quando sistemas de Inteligência Artificial realizarem tratamento de dados pessoais: a) direitos da pessoa afetada por sistema de Inteligência Artificial e os direitos dos titulares; b) a correlação entre sistemas de Inteligência Artificial de alto risco e o tratamento de dados pessoais; e c) mecanismos de governança (ANPD, 2023). O que a autoridade defende é a atração das competências da autoridade competente para ela, como sendo a mais adequada para exercer tal autoridade.

Visto essa possibilidade de convergência entre competências, o interesse da Autoridade Nacional de Proteção de Dados no que se refere à Inteligência Artificial no uso de dados pessoais a torna um dos principais fiscalizadores do uso da tecnologia em território nacional, podendo, dentro das suas competências, intervir frente aos controladores e operadores em defesa do titular de dados, diante da vulnerabilidade que ele e sancionar em caso em que se requeira esse posicionamento. Entretanto, em caso que dados pessoais não são consumidos diretamente por um sistema de Inteligência Artificial, como por exemplo em sistemas industriais que controlam caudeiras ou em aviões quando o sistema auxilia na pilotagem existe, em tese, uma lacuna na fiscalização destes tipos de sistemas.

Neste sentido, a autoridade competente, que segundo o texto de Projeto de Lei 2338, em seu Capítulo VIII, artigo 32, será designada pelo Poder Executivo, restará responsável por regulamentar e fiscalizar o desenvolvimento e o uso de sistemas inteligentes. Em casos em que o sistema de Inteligência Artificial consuma dados pessoais, como principal autoridade para a fiscalização desses tipos de sistemas, espera-se uma fiscalização em conjunto com a

Autoridade Nacional de Proteção de Dados, devido a atração de competência da mesma. Essa autoridade competente será responsável por expedir normas regulatórias, fiscalizar a divulgação de informações exigidas pela lei, fiscalizar o cumprimento da legislação e aplicar sanções em caso de violações.

No esfera consumerista, os Procons possuem protagonismo quanto à fiscalização do uso de sistemas de Inteligência Artificial para fornecimento de produtos e serviços, nas esferas estaduais e municipais, pois o Procon atua na linha de frente na proteção do consumidor e isto não é diferente quanto ao uso de sistemas inteligentes aplicados nesta seara. Destarte, o órgão tem a competência para solicitar explicações e aplicar sanções pelo descumprimento do que está previsto na legislação específica. Por isso, a fiscalização poderia se dar por esse órgão quando a relação de consumo se estabelecer, sem prejuízo à fiscalização por autoridade específica, como a Autoridade Nacional de Proteção de Dados quando houver dados pessoais associados à essa relação de consumo.

O tema também é relevante para o Ministério Público e para o Ministério Público do Trabalho, uma vez que estas entidades têm o papel de fiscalizar os interesses da sociedade e são defensores dos direitos sociais e individuais indisponíveis. O primeiro, inclusive em uma atuação junto ao Poder Judiciário, do qual também faz uso de sistemas inteligentes (Capítulo 5), fiscaliza a fim de assegurar a manutenção da ordem jurídica; o segundo, especificamente em relações trabalhistas, que se utilizam também de sistemas inteligentes em diversos atividades.

Existe, então, a possibilidade de fiscalização por diversos atores. A finalidade é garantir que a legislação seja observada e que direitos e garantias fundamentais sejam respeitados. Contudo, o estabelecimento do tipo de relação pode atrair o protagonismo da fiscalização para uma entidade em específico, mas como há, muitas vezes, certa convergência e confusão em ternos de interesse, mais de uma entidade fiscalizadora poderá atuar com o objetivo de defender o indivíduo e a sociedade.

Portanto, é preciso considerar diversos fatores para concretização do direito à explicação de decisões automatizadas, assim como possíveis soluções para contornar problemas de tecnologia, como a opacidade de sistemas computacionais. Dentre eles, a adoção do nível adequado de transparência, o fornecimento de informações significativas e a possibilidade de revisão humana explicável são instrumentos importantes para evitar prejuízo e danos consideráveis aos titulares de dados e, por consequência, evitar também a responsabilização - discutida Capítulo 3, na Seção 3.3.

5 DO USO DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL E DA MOTIVAÇÃO DE DECISÕES JUDICIAIS

Este capítulo tem como principal objetivo analisar a explicação significativa como instrumento de transparência para a motivação de decisões judiciais a fim de garantir a efetivação do devido processo legal. Com esse propósito, primeiramente, se fez breves reflexões de como a tecnologia é usada nos tribunais brasileiros. Entender o cenário posto e as formas que esse tipo de tecnologia está regrada em âmbito do Poder Judiciário é fundamental para delinear os limites de sua atuação.

Diante da possibilidade em utilizar a Inteligência Artificial para a tomada de decisão em processos judiciais, a argumentação em volta da imparcialidade do tomador de decisão se faz presente. O objetivo consiste em despertar a reflexão de que a atividade da decisão é compartilhada com sistemas automatizados, gerando possíveis decisões que precisam ser explicadas.

Após, discute-se a explicação significativa como um instrumento para concretização do princípio da motivação de decisões judiciais. O intuito desta especificação reside em focar a explicação em um eixo de aplicabilidade como continuidade da discussão sobre explicação de tratamento automatizado por Inteligência Artificial do Capítulo 4. Com isso, delimita-se a discussão em um tipo específico de decisão para análise.

Com esse propósito, a metodologia utilizada conta com uma revisão de literatura narrativa focada nos objetivos deste capítulo. Buscou-se, então, por publicações científicas como artigos científicos nacionais e internacionais, incluindo-se literatura cinzenta, conforme o protocolo no Apêndice B.

5.1 A digitalização dos processos judiciais e seus desdobramentos

A efetivação do Direito por meio de um processo legal adequado é essencial para a resolução de conflitos. Nesse contexto, o Poder Judiciário oferece recursos tecnológicos que, de certa forma, facilitam a utilização, a interoperabilidade, a segurança e proporcionam maior confiabilidade, celeridade e economia ao processo judicial. Embora ainda existam processos físicos, a transição para o ambiente digital é inevitável, e já existem várias plataformas de automação disponíveis.

As plataformas de automação de processos judiciais devem agregar em seu funcionamento os direitos e garantias fundamentais como requisitos de sistema com o objetivo de alcançar o devido processo legal, contemplando: o contraditório, a ampla defesa, a razoável duração do processo; permitindo, deste modo, o acesso à justiça, a igualdade processual, a publicidade dos atos processuais, a motivação das decisões judiciais, dentre outras garantias fundamentais.

Por essa razão, a Constituição detém relação íntima com o processo quando estabelece tais direitos e garantias fundamentais, o controle de constitucionalidade e organiza a

estrutura do Poder Judiciário, assim como também fornece diretrizes para criação de normas infraconstitucionais, das quais estabelecerão, por sua vez, regras específicas para que o juízo possa concretizar o estabelecido na Constituição (MOREIRA, 2014).

O processo judicial adquiriu novos contornos quando se estendeu para o ambiente digital e o Código de Processo Civil possibilitou que atos processuais pudessem ser executados nesse ambiente de forma total ou parcial. O artigo 193 da Seção II do referido código é responsável por permitir a prática eletrônica dos atos processuais, no que se refere a produção, armazenamento, a comunicação e a validação dos atos digitais, deixando para lei específica regrar como se concretizará estas operações e foi a Lei de número 11.419 de 2006, que dispõe sobre a informatização do processo judicial, que coube este papel, sendo ela aplicada aos processos civis, penais, trabalhistas, além dos juizados especiais e em quaisquer outros graus de jurisdição.

Neste sentido, a lei citada no parágrafo anterior permite que os tribunais possam desenvolver sistemas eletrônicos de processamento de ações judiciais, como apontado pelo artigo 8º desde dispositivo legal, não existindo, portanto, a obrigatoriedade de uso de um sistema único para todos tribunais. Embora não haja uma interface computacional de comunicação unificada e padronizada, os sistemas informatizados possibilitam que o operador do direito exerça, de forma efetiva e dentro dos limites estabelecidos pelo software, a produção de atos processuais, incluindo sua eventual alteração. Além disso, por meio desses sistemas, os demais sujeitos processuais também têm acesso ao processo, desde que devidamente autorizados.

Contudo, os órgãos do Poder Judiciário devem, preferencialmente, desenvolver sistemas utilizando programas de código aberto, que estejam acessíveis de forma contínua pela internet e priorizando a padronização. Essa determinação permitiria, em tese, a interoperabilidade entre os sistemas utilizados pelo Judiciário e com padrões aberto a auditoria interna e externa seria igualmente possível.

O Conselho Nacional de Justiça instituiu o sistema de processo judicial eletrônico (Pje) como o sistemas para realização de atos processuais no âmbito do Poder Judiciário, através da Resolução de número 185 de 2013. O artigo 2º do regulamento estabelece que o Processo Judicial Eletrônico abrange o controle do sistema judicial em várias dimensões. Primeiramente, engloba o controle da tramitação do processo, visando garantir uma gestão eficiente e precisa do seu andamento. Além disso, o Processo Judicial Eletrônico atua na padronização de todos os dados e informações que fazem parte do processo judicial, assegurando a uniformidade e consistência desses elementos.

Outro aspecto abordado é a produção, registro e publicidade dos atos processuais, tornando-os transparentes e de fácil acesso para as partes envolvidas e demais interessados. Por fim, o Processo Judicial Eletrônico desempenha um papel importante ao fornecer dados essenciais para a gestão das informações necessárias aos diversos órgãos responsáveis pela supervisão, controle e uso eficaz do sistema judiciário. Essa abordagem busca melhorar a eficiência e a transparência do sistema judicial por meio da automação e da padronização de

processos.

Apesar do Processo Judicial Eletrônico ser a base para automatização do processo em ambiente digital, outros sistemas são utilizados pelos tribunais espalhados pelo Brasil, a exemplo do Sistema de Automatização do Judiciário, conhecido como SAJ, utilizado em estados como São Paulo e Alagoas no âmbito estadual ou o Projudi. Apesar desta variedade de sistemas e a autonomia de escolha de cada tribunal para tanto, nenhum dos sistemas mencionados são *open source*, isto é, não é possível ter acesso ao código fonte das plataformas, permitindo-se apenas uma análise dinâmica de seu funcionamento.

Isso se traduz em uma auditabilidade parcial não sendo possível realizar uma completa, pois não é dada a possibilidade que se possa analisar o que há no código fonte dos sistemas, mesmo que este preste um serviço público. A transparência, neste caso em específico, é diminuída ao mínimo nível, ficando as partes, seus patronos e demais interessados sem ao menos saber como determinada demanda está sendo tratada em seu processo. Não há uma explicação de como ocorre todos os procedimentos dentro dos sistemas, como também não há a ciência do usuário, de como seus dados são armazenados e onde, por exemplo. Direitos e garantias fundamentais relacionado ao processo podem não se efetivar por conta de regras específicas de programação, definidas não pelas partes que se sujeitam ao processo eletrônico, mas por aqueles que, em tese, julgam o processo apenas que transferem as regras da atividade para os projetistas.

Por consequência, os demais sistemas que integram os sistemas judiciais, como os que implementam Inteligência Artificial, também não são abertos. Os códigos fontes, seus modelos de algoritmos, seu dados de treinamento e de teste e sua arquitetura não são passíveis de auditoria por todos. Esse cenário impõe para seus usuários uma lógica de funcionamento que pode ser explicada no modelo *ex ante*, mas não há como apurar se a lógica apresentada realmente condiz com o funcionamento do sistema. Até o momento, parece que o Poder Judiciário não tem dado prioridade à transparência e ao direito à informação, especialmente quando se trata da auditabilidade de sistemas, o que contrasta com os procedimentos rigorosos aplicados ao sistema eleitoral brasileiro e seus componentes.

Essa preocupação, da qual se deveria ter como prioridade, se justifica porque a implementação de sistemas de Inteligência Artificial sem uma consideração adequada de seus impactos e funcionamento pode acarretar sérios riscos para o direito em uma perspectiva sistêmica. Isso porque tais sistemas podem tornar decisões enviesadas inquestionáveis ou, pelo menos, limitar drasticamente o escopo de discussão. Essa abordagem pode resultar em decisões injustas e desiguais, já que as ferramentas de Inteligência Artificial têm o potencial de replicar e amplificar preconceitos e discriminações que existem na sociedade (NUNES; MARQUES, 2018).

Todavia, é possível extrair vantagens de seu uso no judiciário. Segundo Ribeiro e Mazzola (2019), a Inteligência Artificial desempenha um papel fundamental na formação de

precedentes legais, sendo capaz de identificar casos anteriores e até mesmo fazer previsões sobre a probabilidade de sucesso de recursos especiais ou extraordinários. Adicionalmente, outra fonte sugere que robôs podem enriquecer o desenvolvimento do sistema de precedentes, fornecendo estatísticas confiáveis para uma compreensão mais precisa da "jurisprudência dominante". Contudo, é crucial lembrar que a implementação da inteligência artificial no contexto jurídico enfrenta desafios e limitações, como apontado em outra fonte relevante.

Ainda segundo os autores, é de suma importância que a implementação de ferramentas de Inteligência Artificial seja realizada com cautela e após um amplo debate sobre as melhores práticas para o aprendizado de máquina, a fim de minimizar ao máximo qualquer viés nas máquinas. Além disso, é fundamental estabelecer mecanismos que garantam a transparência dos sistemas e dos algoritmos, permitindo uma compreensão adequada do processo de tomada de decisões desses sistemas.

Como efeito de um uso não controlado e transparente desses sistemas, não há garantia de alcance de uma equidade real, ficando, a atuação dos sistemas com Inteligência Artificial, em uma esfera de igualdade formal mais próxima de uma sintaxe das regras contidas no código fonte do próprio sistema. Enquanto a equidade busca considerar as nuances de cada cenário e indivíduo, trazendo um tratamento automatizado que vise incorporar em seu funcionamento as desigualdades e vulnerabilidades das partes envolvidas, fornecendo um senso de justiça computacional; a igualdade no sentido mais formal se aproxima dos sistemas com contextos limitados, como é o caso dos atuais sistemas inteligentes classificados como fracos e que foram concebidos para resolução de problemas específicos, lidando com quantidade também limitada de variáveis mutáveis que não refletem o necessário para balancear as diferenças reais que possam existir.

5.2 Reflexões acerca da utilização da Inteligência Artificial nos tribunais

Os sistemas de Inteligência Artificial representam uma ferramenta singular que pode melhorar a eficácia das atividades do judiciário. De acordo com o trabalho de LAGE (2021), existem dois aspectos da Inteligência Artificial que influenciam o comportamento do juiz: Inteligência Artificial no tribunal e Inteligência Artificial como tribunal. Na primeira abordagem, a Inteligência Artificial pode atuar como uma ferramenta para advogados ou como assistente em processos repetitivos. Já na segunda, a Inteligência Artificial pode ser vista como um substituto para a figura do juiz, uma proposta que não se adéqua à realidade brasileira, ou como um assistente para ajudar o juiz na tomada de decisões.

A Portaria nº 271/2020 do CNJ estabelece as regras para o uso de sistemas de Inteligência Artificial no Poder Judiciário brasileiro. O CNJ é responsável por promover e incentivar investimentos em pesquisa e desenvolvimento de Inteligência Artificial, conforme estipulado no artigo 2º da portaria, que delimita os sistemas inteligentes em três áreas: I) automação de processos judiciais e administrativos, bem como outras atividades do Judiciário; II) análise e

apresentação de dados existentes no Judiciário; III) criação de soluções para apoiar a tomada de decisão dos magistrados ou a elaboração automatizada de minutas de atos judiciais em geral (CNJ, 2020a).

A Resolução nº 332/2020 estabelece diretrizes éticas, transparentes e de governança para o uso da tecnologia de Inteligência Artificial no Poder Judiciário. O artigo 2º da resolução afirma que o objetivo do uso da tecnologia em questão é promover o bem-estar dos jurisdicionados e garantir a prestação equitativa da jurisdição. Além disso, o artigo 4º estipula que os sistemas inteligentes devem respeitar os Direitos Fundamentais. Essas orientações permitem que os tribunais desenvolvam e utilizem ferramentas inteligentes tanto em suas atividades administrativas quanto judiciais (CNJ, 2020b).

Um dos projetos mais significativos é o VICTOR, um sistema de Inteligência Artificial em operação no Supremo Tribunal Federal. Implementado em 2019, o VICTOR está atualmente em fase de produção. Sua principal função é identificar os recursos que incidem em temas mais recorrentes de repercussão geral e devolvê-los aos tribunais de origem. O relatório citado destaca que o principal resultado obtido com o uso do VICTOR foi a redução significativa no tempo de análise dos recursos, diminuindo o tempo médio de 44 minutos para apenas 5 segundos (SALOMÃO et al., 2021).

A ferramenta RAFA 2030 (Redes Artificiais Focadas na Agenda 2030) foi desenvolvida com o propósito de auxiliar na classificação de processos no tribunal, de acordo com os Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 das Nações Unidas. Por meio de redes neurais com análise de simulação semântica, a RAFA 2030 ajuda juízes e servidores a identificar os objetivos presentes em textos de acórdãos ou petições iniciais nos casos do Supremo Tribunal Federal. Essa iniciativa faz parte do projeto estratégico da Agenda 2030, visando transformar o Supremo em uma Corte Constitucional Digital, o que amplia o acesso à justiça e aprimora a transparência dos trabalhos do Tribunal (STF, 2022).

O Supremo Tribunal Federal, até a conclusão desta pesquisa, está finalizando os testes de sua nova ferramenta de Inteligência Artificial, chamada VitórIA. Essa plataforma visa compreender melhor os perfis dos processos no tribunal e permitir o tratamento conjunto de temas similares. A ferramenta identifica automaticamente casos semelhantes no acervo do tribunal, facilitando a identificação ágil e segura de processos aptos a serem tratados em conjunto ou que possam gerar temas de repercussão geral. Isso agilizará a análise e julgamento dos processos, aumentando a eficiência e evitando tratamentos desiguais para casos semelhantes. O

projeto é parte dos esforços do Supremo Tribunal Federal para incorporar a inteligência artificial em suas operações, seguindo projetos anteriores como Victor e Rafa, e marca uma nova etapa na maturidade institucional do tribunal no uso de tecnologias para melhorar os serviços de justiça prestados (STF, 2023).

O sistema inteligente ATHOS, do Superior Tribunal de Justiça (STJ), também é mencionado no mesmo documento. Desde 2019, o sistema está em fase de produção e sua principal funcionalidade é classificar processos e peças agrupando-os por similaridade. O sistema também atua na classificação de acórdãos simuladores, utilizando a jurisprudência do próprio tribunal como base. Além disso, o ATHOS identifica processos com a mesma controvérsia jurídica, com o objetivo de estabelecer teses vinculantes. O relatório citado destaca que o sistema teve resultados positivos, incluindo a redução do número de processos recebidos no âmbito do tribunal, o aumento de recursos representativos da controvérsia e incidentes de resolução de demandas repetitivas, além da uniformização da jurisprudência com o uso de precedentes qualificados (SALOMÃO et al., 2021).

Além do ATHOS, o Superior Tribunal de Justiça (STJ) conta com outros projetos de Inteligência Artificial em desenvolvimento, como o SÓCRATES, E-JURIS e TUA. O SÓCRATES e E-JURIS já estão em fase de produção desde 2019, enquanto o TUA ainda está em fase de desenvolvimento. O SÓCRATES tem como principal objetivo classificar e agrupar processos, bem como identificar precedentes, resultando em uma redução no esforço de triagem e análise de processos, além de auxiliar na seleção de casos representativos da controvérsia. A versão 2.0 do SÓCRATES, que está em fase de desenvolvimento, tem como objetivo otimizar ainda mais a gestão do acervo do tribunal.

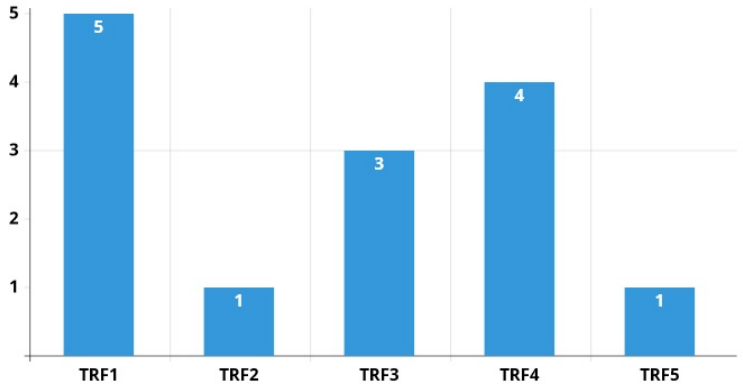
O sistema E-JURIS, por sua vez, é responsável pela extração de referências legislativas e jurisprudenciais citadas nos acórdãos do tribunal, com o objetivo de facilitar o cadastramento das que efetivamente embasaram os votos dos ministros, eliminando as meramente citadas. Isso resulta em maior rapidez no trabalho administrativo da secretaria de jurisprudência. Quanto ao TUA, seu propósito é identificar automaticamente o assunto do processo para distribuição entre as seções do tribunal.

A Plataforma SINAPSES, desenvolvida pelo Conselho Nacional de Justiça e em produção desde 2018, tem como objetivo fornecer uma plataforma para criação e compartilhamento de modelos de Inteligência Artificial entre tribunais. Entre suas principais funcionalidades, destacam-se: a movimentação inteligente, que sugere o movimento a ser aplicado em um despacho; a prevenção, que identifica possíveis casos de prevenção; o gerador de texto jurídico; a triagem de grandes volumes de petições iniciais, classificando-as por temas rotulados previamente; e a verificação de petições, informando se são iniciais ou não.

Existem atualmente 14 projetos de Inteligência Artificial em fase de produção ou desenvolvimento nos Tribunais Regionais Federais. No TRF1, há cinco projetos relacionados, sendo que quatro estão em fase de desenvolvimento - Robô Secor, SIB, ALEI e Projeto Execução Célere - e apenas um em fase de produção - o sistema Banco de Sentenças. No TRF2, apenas

um projeto, chamado ATENDENTE VIRTUAL, está em fase de produção. Já no TRF3, existem três projetos de sistema de Inteligência Artificial, sendo dois em fase de produção - SINARA e SIGMA - e apenas um ainda em fase de desenvolvimento - PREVENÇÃO. No TRF4, há quatro projetos em andamento, sendo que três estão em fase de produção - CLASSIFICAÇÃO DE TEMAS NA VICE-PRESIDÊNCIA E TURMAS RECURSAIS, ANÁLISE DE ASSUNTOS DOS PROCESSOS e TRIAGEM AUTOMÁTICA DE PROCESSOS A PARTIR DA PETIÇÃO INICIAL - e apenas um em fase de desenvolvimento - SUGESTÃO DE MODELOS DE MINUTA. Por fim, no TRF5, apenas um projeto, denominado JULIA - Jurisprudência Laborada com Inteligência Artificial - já se encontra em fase de produção.

Figura 17 – Uso de sistemas de Inteligência Artificial nos TRFs.



Fonte: Baseado em SALOMÃO et al. (2021).

A imagem representada na Figura 17 ilustra de forma gráfica os projetos de Inteligência Artificial em desenvolvimento e em produção nos Tribunais Regionais Federais. Esses sistemas têm funcionalidades diversas, como a extração de dados de documentos, classificação e agrupamento de processos, busca otimizada em documentos e geração de minutas para admissibilidade do juízo.

Os Tribunais Regionais do Trabalho também possuem ou desenvolvem projetos de sistemas com Inteligência Artificial. No TRT1, está em desenvolvimento o projeto “Inteligência Artificial e Eficiência do Judiciário”, que tem como objetivo utilizar análise preditiva em conciliações, sentenças e acórdãos. O TRT4 já possui em produção o sistema “Clusterização de Processos”, que visa agrupar processos semelhantes. Os TRT5, TRT7, TRT15 e TRT20 trabalham em conjunto no projeto “GEMINI”, cujo propósito é classificar e agrupar recursos ordinários similares. No TRT11, o projeto “B.I.TRT 11” está em aprimoramento para incluir funcionalidades que utilizam Inteligência Artificial para acompanhamento de correições e monitoramento de atividades processuais no tribunal. Já o projeto “CONCILIA JT”, no TRT12, é um projeto-piloto que tem como objetivo acelerar o processo de conciliação e identificar casos com potencial para esse tipo de solução.

A tabela apresentada na Figura 18 ilustra a distribuição dos projetos de Inteligência

Artificial nos tribunais superiores, tribunais regionais federais, tribunais regionais do trabalho e tribunais de justiça, classificados em três categorias: em desenvolvimento, em produção e como projeto-piloto. Essa distribuição permite uma visualização clara do avanço da Inteligência Artificial nos diferentes ramos do poder judiciário brasileiro.

Figura 18 – Tabela de distribuição de projetos de Inteligência Artificial.

TRIBUNAIS	EM DESENVOLVIMENTO	PROJETO-PILOTO	EM PRODUÇÃO
Tribunais Superiores (STF, STJ e TST)	4	-	5
Tribunais Regionais Federais	8	2	6
Tribunais Regionais do Trabalho	5	1	1
Tribunais de Justiça	12	4	15
Total por fase	29	7	27

Fonte: Salomão et al. (2021).

Na Figura 18, é possível observar que existem 27 sistemas em fase de produção, dos quais 15 estão na esfera estadual, nos tribunais de justiça. É importante destacar que os tribunais regionais federais apresentam 6 sistemas em fase de produção, enquanto os tribunais regionais do trabalho possuem apenas 1 sistema em fase de produção, o que pode sugerir certa imaturidade nessa esfera. Já os projetos ainda em fase de desenvolvimento somam 29, com a maior quantidade nos tribunais de justiça, contendo 12 projetos. Em seguida, os tribunais regionais federais apresentam 8 projetos em desenvolvimento, enquanto os tribunais regionais do trabalho possuem 5 projetos. Esses dados sugerem que os tribunais regionais federais possuem maior maturidade em relação aos sistemas com Inteligência Artificial em comparação com os tribunais regionais do trabalho.

Por último, o relatório do Centro de Inovação, Administração e Pesquisa do Judiciário - FGV aponta que os sistemas de Inteligência Artificial estão presentes no Poder Judiciário tanto como uma ferramenta para auxiliar atividades administrativas quanto como um auxílio à tomada de decisão - incluindo a geração automatizada de minutas de decisão para aprovação do juiz. É importante ressaltar que, neste último caso, é necessária uma explicação do tratamento automatizado utilizado para apoiar a decisão do juízo. Essa distinção sugere que a Inteligência Artificial está se consolidando como uma ferramenta importante no Poder Judiciário e sua utilização tende a ser cada vez mais explorada e regulamentada.

Independentemente se os sistemas com Inteligência Artificial lidam diretamente ou indiretamente com os processos judiciais ou com processos administrativos, dentro do Poder

Judiciário, desde que exista dados pessoais dentro de seu contexto de análise, e há dados de qualificação pessoal nas peças processuais, os sistemas precisam observar os objetivos dispostos na Lei Geral de Proteção de Dados pessoais, significando dizer que se deve ter um estudo voltado aos impactos do uso da tecnologia em relação ao titular de dados pessoais. Isso deve estar formalizado em um relatório de impacto de proteção de dados complementado com o relatório da análise de impacto algorítmico, realizados periodicamente, visando o monitoramento e controle do uso e dos resultados das plataformas inteligentes. Além disso, sem acesso aos fontes do sistemas e aos dados de treinamento e teste, não é possível analisar com acurácia significativa que represente a realidade de funcionamento dos sistemas, o quão estão adequados ao ecossistema de proteção de dados pessoais.

A análise de adequação desses sistemas se faz importante pela atividade computacional que executam e que se reflete na esfera do direito. Há uma concentração de atividades de categorização/classificação de documentos, a extração de informações, a recuperação de dados, além da geração textual pela maioria dos sistemas de Inteligência Artificial elencados nessa seção. Tais atividades permitem ações como recomendações e a otimização de ações, consequentemente afetando direta e indiretamente os indivíduos, titulares de dados naquele determinado processo ou não.

Diante desse cenário, passa-se a analisar o uso de sistemas com Inteligência Artificial para o auxílio à tomada de decisão do juiz. Isso porque a utilização de sistemas inteligentes não se resume as atividades do juiz, mas pode ser utilizado de forma administrativa com o propósito de otimizar as atividades de gestão.

5.3 A (im)parcialidade das decisões judiciais tomadas com auxílio da Inteligência Artificial

Desde o início, é importante esclarecer que não se deve considerar equivocadamente as decisões tomadas diretamente ou com auxílio de máquinas como neutras ou imparciais, com o objetivo de justificar sua aplicabilidade em processos judiciais. Além de ser inaceitável a perda completa da humanização na administração da justiça, os dados utilizados para alimentar os sistemas de Inteligência Artificial são resultado de interpretações humanas e, portanto, sujeitos à sua qualidade, aos interesses dos projetistas e desenvolvedores e à complexidade do procedimento lógico do algoritmo e da estrutura da arquitetura do sistema. Como consequência, é possível obter decisões excessivamente subjetivas e irregulares, resultando em algo conhecido como “algoritmos enviesados” (ROQUE; SANTOS, 2021).

Ainda para Roque e Santos (2021), esses algoritmos enviesados podem trazer danos irreparáveis, já que apresentam padrões distorcidos em sua formação, afetando o funcionamento do sistema. Apesar de aparentarem neutralidade, eles perpetuam vieses difíceis de serem identificados, até mesmo pelos seus próprios desenvolvedores, por falta de transparência. Isso pode levar à legitimação, através da tecnologia, de tratamentos desiguais e discriminatórios que

já existem na sociedade e são assimilados pelos algoritmos.

No contexto jurídico, por exemplo, cita-se o sistema *Correctional Offender Management Profiling for Alternative Sanctions* - COMPAS, utilizado nos Estados Unidos para avaliar o risco de reincidência e auxiliar na fixação da pena do acusado. O sistema possuía um viés social com resultado discriminatório, indicando que os acusados negros eram mais propensos à reincidência do que os acusados brancos, e não garantia ao acusado acesso aos dados relacionados ao processo computacional que levou a essa conclusão (BRENNAN; DIETERICH, 2018).

Isso leva, indubitavelmente, para um cenário de vulnerabilidade tecnológica da qual atinge aqueles que são submetidos a este tipo de tratamento automatizado. Os direitos e garantias fundamentais ligadas ao processo podem ser desprezados e ignorados por parte dos responsáveis pelo desenvolvimento do sistema, dos quais podem transportar para o software os equívocos e preconceitos modelados no projeto.

Nessa perspectiva, a imparcialidade do julgador pode restar invalidada, porque não há garantia que as partes envolvidas tenham um julgamento justo e isento de influências externas que possam comprometer a equidade do processo. Como as regras da atividade estão ocultas nos códigos do sistema, o julgador que meramente acatar sem ao menos realizar uma análise mais aprofundada da decisão sugerida não age de forma neutra o suficiente para estabelecer um distanciamento de favorecimento para qualquer uma das partes.

A imparcialidade no julgamento deve ser avaliada tanto objetivamente quanto subjetivamente. A avaliação objetiva considera a conduta do juiz, levando em conta seu comportamento e as circunstâncias do processo. Já a avaliação subjetiva leva em consideração a percepção que as partes envolvidas têm em relação ao julgamento. Isso afeta diretamente a avaliação subjetiva da imparcialidade, pois as partes podem perceber a imparcialidade ou parcialidade do julgador com base em sua conduta.

Essa imparcialidade em julgar consiste em um princípio importante para o indivíduo que recorre ao Poder Judiciário em busca da solução de determinado problema que não consegue resolver por forças próprias. No Brasil, essa garantia está normatizada em diversos instrumentos legais, como o Código de Ética da Magistratura, o Código de Processo Civil, a Lei Orgânica da Magistratura e a Constituição Federal.

O Código de Ética da Magistratura, CNJ (2008), estabelece que a conduta do juiz deve ser pautada pelos princípios da independência, imparcialidade, conhecimento e capacitação, cortesia, transparência, segredo profissional, prudência, diligência, integridade profissional e pessoal, dignidade, honra e decoro. O Código de Processo Civil, Brasil (2015), reforça o princípio da imparcialidade ao determinar que o juiz deve assegurar às partes igualdade de tratamento e se afastar do caso em que tiver algum envolvimento pessoal na causa. Já o Código de Processo Penal, Brasil (1941), em seu artigo 254, estabelece diversas hipóteses de suspeição do juiz, como ser amigo íntimo ou inimigo capital de qualquer uma das partes ou ter aconselhado qualquer uma delas.

A Lei Orgânica da Magistratura, Brasil (1979), por sua vez, define as atribuições do juiz em cumprir e fazer cumprir as disposições legais e os atos de ofício com independência, serenidade e exatidão. E ainda veda que o juiz manifeste opinião sobre processo pendente de julgamento, seu ou de outrem, ou juízo depreciativo sobre despachos, votos ou sentenças de órgãos judiciais.

A Constituição Federal de 1988, por sua vez, garante o devido processo legal, que inclui a garantia da imparcialidade do juiz. Assim, fica claro que a imparcialidade é um dos direitos processuais da parte e deve ser respeitada para assegurar a justiça e a equidade no julgamento das causas. O juiz deve conduzir o processo sem se deixar influenciar por interesses pessoais, política, econômicos ou sociais.

Na esfera internacional, o Pacto Internacional dos Direitos Civis e Políticos, Organização das Nações Unidas, ONU (1966), é um tratado internacional adotado pela Assembleia Geral das Nações Unidas em 1966, que entrou em vigor em 1976. O objetivo deste pacto é proteger e garantir os direitos civis e políticos de todos os indivíduos, incluindo o direito a um julgamento justo e imparcial. O artigo 14 do pacto estabelece o direito a um julgamento justo, que inclui o direito a um tribunal imparcial e independente. Isso significa que a pessoa que está sendo julgada deve ter a garantia de que o tribunal não tem qualquer predisposição ou preconceito a favor ou contra ela e que o tribunal não esteja sujeito a influências externas, políticas ou econômicas que possam afetar sua decisão.

Há uma série de instrumentos normativos que atribuem uma obrigação legal para que se exista um julgamento livre de interferências tendenciosas que ao final entregue a jurisdição de forma justa. O juiz, mesmo que não trabalhe ativamente na geração de uma minuta de decisão e delegue isto para uma máquina, ainda é responsável pelo resultado final, ou seja, a decisão. É ele quem chancelará a minuta e a converterá em decisão judicial, existindo uma atividade supervisionada do sistema.

Portanto, quando se confirma a minuta gerada pelo sistema de Inteligência Artificial, a parcialidade ou imparcialidade do próprio sistema será absorvida pela pessoa do juiz que a assumirá como dele. Em razão disso, as decisões tomadas com auxílio de sistemas inteligentes devem ser motivadas, assim como as decisões criadas pela pessoa do juiz. Através desta motivação, não apenas será possível saber quais os critérios adotados para a decisão em si, mas também será possível avaliar se o juiz é parcial ou imparcial no ato. O mesmo deve ocorrer em cede de recurso, quando a decisão proferida será reexaminada também por um sistema inteligente do qual sugerirá um desfecho. Repetindo o que ocorrera na origem do processo, o elemento humano servirá como fiscalizador e chancelador do que o sistema automatizado fará.

De qualquer forma, a imparcialidade do julgador não pode ser delegada ao software. Ainda mais porque a falta de transparência contida nesses sistemas atrapalha ou inviabiliza a devida auditoria sobre seu funcionamento. Por esta razão que a explicação significativa do funcionamento da Inteligência Artificial é essencial. Ela traz novamente ao processo o caráter

humanizado, uma vez que possibilita a ação apropriada das partes e que o julgador possa analisar o que foi feito para assumir a decisão como sua.

5.4 A explicação significativa como instrumento de concretização da motivação de decisões judiciais

Quando se considera a tecnologia de Inteligência Artificial para auxílio à tomada de decisão, espera-se que o produto resultante do processo automatizado seja uma minuta de decisão, isto é, um texto estruturado composto pelos elementos necessários de uma decisão judicial. Esse artefato é constituído substancialmente por linguagem de alto nível (linguagem compreensível pelo ser humano) através de processamento de linguagem natural, área especializada da Inteligência Artificial, como apresentado na Capítulo 2, na Subseção 2.3.2.

Já a forma de construção dos sistemas que lidam com processamento de linguagem natural, dentre as diversas técnicas que podem ser utilizadas, a abordagem de aprendizado profunda, montados com base nas redes neurais artificiais profundas (modelo mostrado como exemplo na Figura 16), se destaca. Esse modelo possibilita trabalhar com volumes significativos de dados, identificando padrões nos textos e gerando saídas eficientes.

Além disso, outras técnicas podem trabalhar em conjunto, tais como: a Tokenização, divisão de um texto em unidades menores (*tokens*) como palavras, frases ou sentenças; a análise morfológica, identificação da estrutura gramatical das palavras, incluindo a sua classe gramatical (substantivo, verbo, adjetivo, etc.) e suas flexões (gênero, número, tempo, etc.); a análise sintática, análise da estrutura gramatical das frases, incluindo a identificação das suas partes constituintes (sujeito, verbo, objeto, etc.) e suas relações hierárquicas; Reconhecimento de entidades nomeadas (NER), identificação de entidades nomeadas em um texto, como nomes de pessoas, organizações, lugares, datas, etc; a análise semântica, análise do significado do texto, incluindo a identificação de relações entre as palavras e frases; e a modelagem de tópicos, técnica utilizada para identificar tópicos e temas em um conjunto de documentos (JURAFSKY; MARTIN, 2019).

Todas as técnicas utilizadas têm como finalidade gerar comunicação com o ser humano e podem ser empregadas para geração das minutas de decisões judiciais que serão sugeridas ao juízo, que pode acatá-la ou não. Para que essa minuta possa ser acatada pelo juízo, no entanto, ela deve ser coerente com o processo judicial, portanto, cada palavra escolhida pelo sistema e incluída na minuta detém importância, pois é através dela que a razão de ter indicado uma decisão se justifica.

A razão de decidir é um dos pontos mais importantes na decisão judicial, principalmente ao se tratar de sentenças e acórdãos judiciais, pois permite conhecer como o julgador atingiu determinada lógica de conclusão. Essa razão de decidir se torna a motivação da própria decisão judicial e, com isto, acrescenta à decisão a conexão com os procedimentos, pedidos e argumentos realizados pelas partes.

No âmbito jurídico, a interação entre racionalidade e justificação de decisões pode ser conceituada considerando que as decisões judiciais devem estar embasadas em razões que possam ser escrutinadas e sujeitas a revisão. A justificação reflete o grau de racionalidade presente na decisão, e, para estabelecer uma confiança nas instituições, é essencial que as decisões sejam adequadamente fundamentadas (MONTEIRO, 2007).

Naturalmente, ainda segundo Monteiro (2007), uma decisão judicial segue critérios *a posteriori* e sua justificação visa torná-la aceitável. Entretanto, em certos casos, influências significativas externas podem afetar a aplicação do Direito, resultando em uma situação em que a tarefa da justificação passa a ser meramente uma tentativa de racionalizar a posteriori uma decisão tomada *a priori*, sob a influência de fatores emocionais, e ainda apresentá-la como razoável. Nessas circunstâncias específicas, pode não se dá uma ênfase substancial à consistência das razões expressas em seu fundamento e praticamente qualquer argumento que aparente ter lógica é usado para justificar a decisão de forma persuasiva. Isso se torna ainda mais relevante quando utilizada tecnologia de automatização de tarefas, como a geração de minuta para decisão judicial.

A motivação das decisões judiciais é um direito fundamental garantido pela Constituição Federal, expresso no artigo 93, IX. Esse direito está intimamente relacionado ao princípio do devido processo legal e à busca pela justiça. É fundamental que as partes envolvidas saibam o motivo exato de sua vitória ou derrota, e possam avaliar se a decisão proferida foi correta, o que lhes permite recorrer adequadamente (LUCCA, 2015). Além disso, a motivação das decisões judiciais é uma forma de prestação de contas da atividade jurisdicional, tanto para as partes envolvidas quanto para a sociedade como um todo.

O dever de motivação das decisões judiciais, como discutido por Lucca (2015), desempenha um papel fundamental na garantia de diversos direitos processuais, tais como o contraditório e a ampla defesa, no sentido de permitir que as partes possam apresentar contrarrazões fundamentadas e com base em argumentos sólidos em contraponto as rações apresentadas pelo juiz; a inércia jurisdicional, o que justifica as decisões judiciais, seja para deferir ou indeferir um pedido, garantindo que a jurisdição seja exercida de forma responsável e fundamentada; o princípio dispositivo, impondo ao juiz o dever de se ater aos limites do pedido formulado pelas partes, além de trazer para o processo o tratamento isonômico entre elas; o duplo grau de jurisdição, permitindo o reexame das decisões judiciais e isto ocorre a partir da motivações das decisões; a imparcialidade do julgador, impondo, igualmente, ao juiz o dever de fundamentar suas decisões com base em argumentos objetivos e racionais, evitando que decisões sejam tomadas com base em preconceitos, interesses pessoais ou outros fatores subjetivos; dentre outros direitos relevantes.

Por essa importância, portanto, que as decisões judiciais são compostas por três elementos, conforme previsto no artigo 489 do Código de Processo Civil: 1) o relatório, que apresenta os pontos relevantes do processo até então; 2) os fundamentos, que incluem a análise das questões de fato e de direito; e 3) o dispositivo, que traz a resolução das questões principais

submetidas pelas partes para análise. Portanto, uma decisão judicial genérica que não aborde todos os pontos levantados como argumentos capazes de levar a uma conclusão ou que se limite a indicar ou reproduzir normas sem explicar adequadamente o motivo de sua aplicação, não estará fundamentada nos termos do artigo 489, §1º. Além disso, uma decisão que use conceitos jurídicos indeterminados ou não atenda a outros parâmetros estabelecidos também não estará devidamente fundamentada, sendo nula em decorrência dessa falta de fundamentação.

Na esfera administrativa ocorre o mesmo entendimento com a falta de motivação de uma decisão administrativa automatizada. Pode-se citar o Enunciado 12 da I Jornada de Direito Administrativo, organizada pelo Conselho da Justiça Federal, vinculado ao Superior Tribunal de Justiça, que diz que a decisão administrativa tomada por robôs deve ser devidamente fundamentada, e sua falta de transparência pode levar à sua anulação (CJF, 2020).

O processo decisório no âmbito judicial envolve diversas etapas complexas, que incluem a análise e interpretação de dados coletados ao longo do processo, como provas documentais, periciais, testemunhais, entre outras, o que é fundamental para identificar as informações relevantes para a tomada de decisão. É importante ressaltar que todos os elementos levados aos autos do processo devem ser considerados no momento da decisão, incluindo-se os elementos de prova, independentemente de quem os tenham levado e, por força do artigo 371 do Código de Processo Civil, na decisão, o juiz apontará de forma clara e explícita as bases que fundamentam sua convicção e qualquer software que se proponha a apoiar o juiz deve ser capaz de ponderar todos os pontos apresentados na motivação (LEONARDO; ESTEVÃO, 2020). Embora não seja o objetivo principal deste trabalho a análise de provas por sistemas com Inteligência Artificial, cabe realizar algumas considerações sobre isso.

Primeiramente, cabe evidenciar que, segundo Didier Jr, Braga e Oliveira (2015), as provas podem ser classificadas, quanto ao objetivo por: I) provas diretas são aquelas que fornecem uma conexão evidente e imediata com o fato que se busca provar; e II) provas indiretas, também conhecidas como provas circunstanciais, não estabelecem uma conexão direta com o fato, mas fornecem indícios ou evidências que, quando combinadas, levam a uma conclusão sobre o que ocorreu. A Inteligência Artificial pode ser um instrumento eficiente na análise de provas, quando se trata de provas diretas, desde que se considere os vários tipos de forma em sua construção e finalidade. Já para as provas indiretas, A Inteligência Artificial pode auxiliar na organização e categorização de informações, mas a avaliação do significado e do peso dessas evidências ainda requer a intervenção humana, especialmente em casos complexos que envolvem inferências e raciocínio lógico.

Analisar os artefatos que reduzem a prova em uma linguagem natural dentro do processo, como a redução à termo do relato de testemunhas ou declarantes e mesmo provas documentais ou relatórios periciais, é algo mais tangível de ser fazer com um sistema que já lida com o processamento de linguagem natural de forma escrita. Porém, deve-se levar em consideração os diversos tipos de formatos de provas que podem adentrar em um processo, como vídeos, áudios e fotografias, por exemplo. Neste sentido, o sistema deverá incorporar o processamento

de sinais digitais, capazes de lidar com dados não estruturados em formato de mídia digital e processamento de imagens digitais, capaz de tratar imagens de diversos tipos e formatos.

Nesse aspecto, para que um sistema inteligente consiga realizar qualquer processamento nos diversos formatos possíveis, deve ele ter sido planejado para tal. Não se pode esperar que um sistema com Inteligência Artificial faça alguma tarefa que esteja contrário ao seu domínio, mesmo ele tendo certo grau de inovação sem uma programação explicita, isto é, o sistema pode criar certos caminhos que o leva a resultados novos, mas isto não o torna resiliente o suficiente para lidar com todo e qualquer cenário que lhe for apresentado.

Ponto de interesse se refere às provas indiciárias, que no direito civil substancial caracteriza-se como presunção, como disposto pelo artigo 212, inciso IV do Código Civil, pois há uma presunção do uso da lógica utilizando-se do método indutivo e isso poderia resultar em uma conclusão de que tais provas teriam mais aceitação pelos sistemas utilizados, mesmo sendo caracterizada com uma prova indireta. Apesar dos sistemas lidarem com a lógica, ela não necessariamente é uma lógica clássica, da qual se lida com proposições através de deduções e induções a partir da linguagem.

A lógica que o sistema possuí é a da programação dentro do seu código, que se lida, nos modelos mais atuais de Inteligência Artificial, com um fluxo de instruções derivados de algoritmos não lineares, isto é, instruções que não seguem uma simples lógica de se-então-senão. Além disto, para que exista tal aplicação lógica, é preciso que o sistema considere o que analisa como suficiente ou não para apontamento de uma veracidade, no mínimo, dos fatos e não somente uma análise de verdadeiro ou falso.

Independentemente do formato a prova, ela deve existir como representação de entrada para um sistema computacional processar. Isso quer dizer que o sistema deve estar preparado para receber tal representação e processar uma saída significativa. Igualmente, caso exista uma implementação neste sentido, deve existir uma explicação significativa pormenorizada do como o tratamento dessas provas submetidas ao sistema foi processada. Essa explicação permitirá o enfrentamento da prova em termos de contraditório e ampla defesa. Além disto, é imprescindível a existência de revisão humana como alternativa para a análise computacional automatizada, caso o contrário, a prova e todos os atos processuais decorrentes dela deverão ser consideradas nulas.

Uma das principais dificuldades no uso de tecnologias de software, tanto em programas construídos de forma tradicional quanto em técnicas avançadas de Inteligência Artificial, está no fato de que os modelos de funcionamento utilizados são predefinidos pelo ser humano, o que se traduz como algoritmos e dados. Esses modelos possuem níveis de transparência variados em relação ao entendimento da lógica de execução, ou seja, a depender do modelo, a previsibilidade de sua resposta se torna inexistente. Assim, a transparência se mostra necessária pois se relaciona diretamente com os preceitos do princípio da motivação, permitindo a prestação de contas já citada nessa seção.

Por essa limitação da atuação do sistema de Inteligência Artificial, quando ele trata

principalmente de linguagem natural, deve-se ter cuidado com a geração automatizada do textos. Isso porque, como defende Warat (1995), a linguagem jurídica é uma criação humana, construída socialmente e que, portanto, é uma ferramenta para a compreensão e criação da realidade jurídica. Isso é algo que um sistema puramente artificial não consegue alcançar por conta de sua natureza limita.

Além do mais, segundo Eco et al. (1993), a interpretação de um texto envolve a tarefa de compreender o porquê de determinadas palavras e frases possuírem diferentes significados e sentidos, enquanto outras não. É necessário analisar cuidadosamente a maneira como esses elementos são interpretados e como se relacionam uns com os outros dentro do contexto do texto. Isso requer habilidade em decodificar a linguagem utilizada, mas também uma compreensão mais profunda das ideias e conceitos que estão sendo expressos. Em outras palavras, a interpretação de um texto envolve tanto uma análise linguística quanto uma análise crítica do conteúdo, para que se possa explicar adequadamente o significado por trás das palavras.

Então, para formação de uma minuta decisória, construída com base na tecnologia de processamento de linguagem natural, a explicação do porquê um sistema computacional inteligente escolheu certas palavras em detrimento de outras também é um ponto de enfoque importante para explicabilidade de uma decisão tomada com auxílio desses sistemas. Neste sentido, Nogueira (2009), aborda os conceitos de interpretação-conhecimento e interpretação-decisão. A primeira refere-se a um processo de compreensão ou descrição do significado de uma expressão específica. Já a interpretação-decisão consiste em atribuir um significado específico a uma expressão, escolhendo uma interpretação em detrimento de outras possíveis. Esse cenário remete ao questionamento de como um sistema com implementação de tecnologia de Inteligência Artificial considerou as entradas textuais para produzir seu texto de saída.

Não obstante a todos os desafios que um sistema inteligente deve enfrentar em relação à linguagem e a semântica, assim como os requisitos legais para motivação. As provas, dentro do processo, também constituem fator de impacto para a geração de minuta de decisão. Isso porque, segundo Mitidiero (2020), a regra do ônus da prova, nesse sentido, serve como um critério de desempate, atribuindo a responsabilidade da prova para uma das partes envolvidas no processo. Essa regra tem como objetivo trazer racionalidade, objetividade e segurança jurídica para o processo decisório, garantindo que a decisão seja baseada em provas sólidas e não em meras alegações. Portanto, o julgamento conforme o ônus da prova é um dever judicial diante do fato duvidoso, uma vez que sua aplicação é essencial para garantir a justiça e a imparcialidade no processo decisório. Explicar, portanto, quais critérios foram adotados pelo sistema inteligente para pesar quais provas são relevantes ou não e como elas influenciaram a proposta de decisão torna-se fundamental para a lisura do processo.

Neste ponto, a transparência possui dois aspectos relevantes dentro do processo, o primeiro se relaciona com os níveis de transparência apresentados na Seção 4.1 do Capítulo 4, direcionado à explicação significativa através de características específicas e o segundo se

relaciona com a transparência da tecnologia utilizada nos sistemas de Inteligência Artificial, por exemplo. A transparência direcionadora da explicação significativa, que a priori, traz consigo o nível de transparência total, pois o processo judicial precisa ser publicizado e acessível a qualquer pessoa que deseje consultá-lo, com exceção daqueles que estivem sobre sigilo ou segredo de justiça, é afetada diretamente pela transparência ou sua falta ao se tratar das tecnologias digitais utilizadas, como os sistemas de Inteligência Artificial.

Embora esses sistemas possam apresentar um alto grau de acurácia em sua atividade, é importante destacar que ainda assim podem ocorrer resultados incorretos ou discriminatórios. Isso porque a tomada de decisão é uma atividade complexa que envolve diversas outras atividades igualmente complexas, e nenhum sistema está imune a erros e falhas. Razão essa que a transparência tecnológica é fator determinante para o uso da tecnologia no âmbito do Poder Judiciário no Brasil.

Diante disso, a publicização algorítmica, que é um processo que envolve a disponibilização pública dos algoritmos utilizados em decisões judiciais, é de vital relevância. Esse processo permite que as partes envolvidas possam entender como as decisões foram tomadas e questionar a validade dos resultados. Desta forma, ao tornar o processo algorítmico transparente, é possível verificar se há vícios ou tendenciosidades que possam comprometer a idoneidade do sistema. Para além disso, segundo Kim (2017), essa auditoria pode revelar quando ele exclui inadvertidamente grupos protegidos, permitindo ao usuário se engajar em autoexame e revisar seus processos para eliminar vieses implícitos ou não intencionais. A publicização algorítmica, portanto, é vista como um caminho viável para garantir que o sistema judicial funcione de maneira justa e equitativa. Assim, ao conhecer como o sistema funciona ou funcionará para gerar as minutas submetidas à análise do magistrado, é possível identificar se há preconceitos ou discriminações que possam estar embutidos nos algoritmos utilizados. Por isso, essa publicização dos algoritmos pode ajudar a evitar decisões injustas ou discriminatórias.

Para Vale (2019), dentro deste contexto, alguns especialistas defendem que a transparência processual prevista no artigo 5º, inciso LX da Constituição Federal e no artigo 8º do Código de Processo Civil, precisa ser ampliada para incluir a publicidade algorítmica. Em outras palavras, é necessário que sejam divulgados detalhes sobre os algoritmos utilizados para tomar decisões. Ao incluir a publicidade algorítmica na transparência processual, seria possível garantir a integridade e a equidade nas decisões judiciais, ampliando a compreensão e a aceitação do sistema. Portanto, a atualização dos contornos da publicidade processual para incluir a publicidade algorítmica é vista como uma medida necessária para garantir a justiça no sistema judicial. No mesmo sentido, Goodman (2016), traz a ligação entre essa transparência vinda da publicização algorítmica e a explicação quando afirma que o princípio da transparência do algoritmo estabelece o "direito à explicação", que confere aos indivíduos titularidade dos dados o direito a receber "informações relevantes sobre o processo lógico utilizado, bem como o significado e as possíveis consequências" no caso de decisões automatizadas ou criação de perfil.

Embora o termo publicização algorítmica se refira ao funcionamento de todo o sistema de software, com Inteligência Artificial ou não, o mais apropriado, para a devida observação à transparência e por consequência ao princípio da motivação aqui discutido, seria publicização da arquitetura do sistema, pois além dos algoritmos, a representação de conhecimento (o como os dados estão organizados dentro da base de conhecimento) e os requisitos da atividade se encaixam para a execução do sistema automatizado, o processo de construção de software e sistema foi apresentado de forma resumida no Capítulo 2.

Todavia, a publicização dos algoritmos e da arquitetura do sistema pode representar um ponto negativo na utilização desses tipos de software. O fato de se conhecer o como esses sistemas funcionam torna possível agir para para direcionar o resultado de saída dele. Não há garantias de que não se possa manipular o seu funcionamento com objetivo de obter vantagens perante a parte adversária. Porém, ainda sim, a transparência traz vantagens consideráveis em comparação as suas desvantagens.

Destaca-se que a publicização dos algoritmos e da arquitetura dos sistemas não significa a vedação ao sigilo dos atos processuais, o que iria de encontro ao artigo 5º, inciso LX, do qual diz que a lei somente poderá limitar a divulgação dos atos processuais quando houver a necessidade de proteger a privacidade ou atender ao interesse público, em conjunto com o artigo 93, inciso IX, que também diz que a lei pode restringir os atos processuais, limitando-os às partes envolvidas e seus advogados, ou apenas a estes, quando necessário para proteger o direito à privacidade, desde que isso não prejudique o interesse público à informação. O código de Processo Civil, em seu artigo 189, dispõe os casos em que o segredo de justiça deve ser aplicado. O código estabelece que em certos processos, o segredo de justiça deve ser sempre mantido, mas também permite que seja decretado quando houver um interesse público nisso. Portanto a lei restringirá o acesso aos atos processuais, assim como obstar o acesso à fonte da tecnologia, visando proteger a livre iniciativa e os segredos industriais e comerciais. Contudo, como existe interesse público relevante para a publicização da arquitetura do sistema utilizado, parece razoável que se integre a transparência processual tal arquitetura de sistema.

Por outro lado, defende Malgieri e Comandé (2017) que uma decisão judicial deve ou não ser explicada e como isso deve ocorrer varia com base em vários fatores, incluindo a importância da decisão e a natureza do tomador de decisão. Ainda argumenta que, por exemplo, um juiz que decide sobre uma petição para conceder uma audiência geralmente pode fazê-lo com pouca ou nenhuma explicação, o que seria uma decisão essencialmente discricionária. Mas um juiz que dá uma sentença criminal - uma das decisões mais importantes que o judiciário pode tomar - deve fornecer uma explicação para que o réu possa detectar e desafiar qualquer impropriedade ou erro. Os autores levam em consideração o grau de importância da decisão para que esta tenha igualmente um nível de explicação condizente com a decisão.

Contudo, mesmo que a decisão tenha menor grau de importância, ela deve ser explicada em sua motivação, mesmo que seja para marcação de uma audiência. Por conta de sua natureza menos complexa, uma decisão como esta pode ser explicada com facilidade em comparação

com outras decisões, levando-se em consideração o tipo de modelo utilizado.

Todas as decisões judiciais devem ser motivadas, expressando os motivos de fato e de direito que levaram ao convencimento do magistrado, conforme previsto na Constituição (GILLET; PORTELA, 2018). Isso se aplica inclusive às decisões apoiadas por sistemas de Inteligência Artificial. Portanto, é importante fornecer uma explicação significativa que abrange tanto a explicação *ex ante* quanto a *ex post*, ou seja, todos os aspectos voltados ao tratamento automatizado que resultou na decisão. Isso inclui informações sobre o uso de sistemas inteligentes na geração da minuta da decisão, bem como a explanação sobre o modelo do sistema e como cada dado inserido impactou no resultado gerado pelo sistema, que foi chancelado pelo juiz.

5.5 As perspectivas futuras da utilização de sistemas de Inteligência Artificial no Poder Judiciário

Retomando o relatório o relatório do Centro de Inovação, Administração e Pesquisa do Judiciário - FGV, Salomão et al. (2021) - a Inteligência Artificial está presente nos tribunais brasileiros e esse caminho se manterá devido a melhora dos procedimentos administrativos e judiciais da qual ela é aplicada. Essa Seção tem como objetivo apresentar e discutir as perspectivas futuras para o uso da tecnologia de Inteligência Artificial dentro dos tribunais brasileiro, incluindo-se o aspecto do auxílio à tomada de decisão.

Com base no relatório apresentado na Seção 5.2, o uso da tecnologia será utilizada em diversas atividades dentro do Poder judiciário brasileiro e algumas vertentes tem maior potencial, são elas: processamento de grandes volumes de dados/informações, utilizada para ajudar os juízes a lidar com grandes quantidades de informações, como processos e documentos, e fornecer *insights* para a tomada de decisão; análise preditiva, usada para prever decisões judiciais com base em padrões e histórico de casos, ajudando os juízes a tomar decisões mais informadas; automação de processos, usada para automatizar tarefas repetitivas, como a organização de documentos, permitindo que os juízes se concentrem em tarefas mais importantes; identificação de casos semelhantes, pode ajudar a identificar casos semelhantes e fornecer informações sobre como esses casos foram tratados no passado, permitindo que os juízes tomem decisões mais consistentes; e melhoria da eficiência, pode ajudar a melhorar a eficiência do sistema judicial, permitindo que os juízes tomem decisões mais rapidamente e reduzindo a carga de trabalho em geral.

Quanto ao primeiro caso, de processamento de grandes volumes de dados/informação, o Poder Judiciário brasileiro lida com um grande volume de informações e processos em andamento, o que pode sobrecarregar os juízes e dificultar a análise de dados relevantes para a tomada de decisão. Nesse sentido, a inteligência artificial pode ser uma ferramenta valiosa para auxiliar os juízes na análise de grandes quantidades de informações e documentos, como processos e jurisprudências, por exemplo.

A utilização de algoritmos de aprendizado de máquina permite que a Inteligência Artificial identifique padrões e relações em dados complexos, fornecendo precisão e relevância para os juízes tomarem decisões informadas. Essa abordagem pode ajudar a reduzir a carga de trabalho dos juízes, permitindo que eles concentrem seus esforços em casos mais complexos e exigentes.

A análise preditiva é uma das aplicações da Inteligência Artificial no Poder Judiciário. Através do uso da aprendizado de máquina, a Inteligência Artificial pode examinar grandes quantidades de dados históricos de casos judiciais, identificar padrões e tendências em como os juízes tomaram decisões em situações semelhantes no passado. Com base nesses dados, a Inteligência Artificial pode então prever o resultado mais provável de um caso em particular, fornecendo informações valiosas para os juízes em sua tomada de decisão. Isso não significa que a Inteligência Artificial tomará a decisão em si, mas sim que fornecerá aos juízes uma avaliação preditiva das possibilidades de resultado do caso, permitindo que eles tomem uma decisão mais informada e fundamentada. Essa abordagem pode ajudar a reduzir o risco de decisões inconsistentes e injustas, além de economizar tempo e recursos para o Judiciário.

No entanto, é importante observar que a análise preditiva é apenas uma ferramenta e não deve substituir a expertise e julgamento humano dos juízes. Ainda é necessário que os juízes apliquem seu conhecimento jurídico e avaliem cada caso individualmente antes de tomar uma decisão final.

A utilização de Inteligência Artificial no Poder Judiciário também pode se estender para a automação de processos, especialmente em tarefas repetitivas e de baixo valor agregado. Com o auxílio de técnicas de processamento de linguagem natural e visão computacional, a Inteligência Artificial pode ser usada para organizar e classificar documentos, extrair informações relevantes, preencher formulários e até mesmo redigir decisões padrão. Essas tarefas podem ser realizadas de forma mais rápida e eficiente do que se fossem realizadas manualmente, permitindo que os juízes se concentrem em tarefas mais complexas e críticas para a tomada de decisão. Além disso, a automação de processos também pode levar a uma maior uniformidade e consistência nas decisões, reduzindo o risco de erros e garantindo uma maior previsibilidade para as partes envolvidas no processo judicial.

Identificar casos semelhantes por meio da análise de seus dados e metadados, incluindo informações sobre partes envolvidas, fatos, argumentos e decisões anteriores. Essa identificação pode ajudar os juízes a entender como casos semelhantes foram tratados no passado e a aplicar esses precedentes à tomada de decisão em novos casos. Além disso, a Inteligência Artificial pode ser utilizada para sugerir quais precedentes e jurisprudências são mais relevantes para o caso em questão, permitindo que os juízes tomem decisões mais informadas e consistentes. Essa abordagem pode ajudar a reduzir a inconsistência na jurisprudência e melhorar a previsibilidade das decisões judiciais.

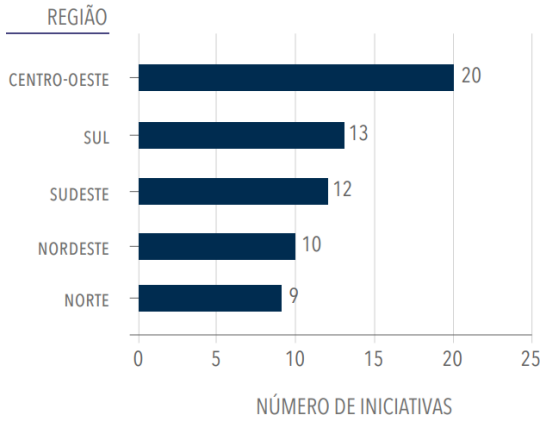
Todas as atividades citadas nesta seção podem trazer eficiência ao processo judicial, influenciando a duração do processo, otimizando os procedimentos até a entrega de jurisdição e

encerramento do processo. Por essa razão, a tecnologia em análise contribui para a tangibilidade do princípio do tempo razoável do processo, do qual diz que todas as pessoas, tanto no âmbito judicial quanto administrativo, têm o direito de que seus processos sejam concluídos em um tempo razoável e que sejam implementados meios que garantam a rapidez em sua tramitação, como consta o artigo 5º, LXXVIII, da Constituição Federal. Isso porque a demora excessiva no andamento do processo pode prejudicar os direitos das partes envolvidas, além de causar insegurança jurídica e afetar a credibilidade do sistema judicial.

A Inteligência Artificial também pode contribuir para assegurar o cumprimento do princípio do tempo razoável do processo ao identificar gargalos no sistema judicial. Por meio da análise dos fluxos de trabalho, a tecnologia pode detectar processos que apresentam maior lentidão do que o esperado, possibilitando a tomada de medidas para solucionar esses problemas e, assim, acelerar a tramitação processual. Dessa forma, a utilização desses sistemas pode promover uma gestão mais eficiente dos processos judiciais e garantir a observância do direito fundamental à razoável duração do processo.

Essa eficiência já foi percebida pelos tribunais em todo o Brasil, como mostra a Figura 19. Em todas as regiões do país encontram-se iniciativas que abrangem a tecnologia.

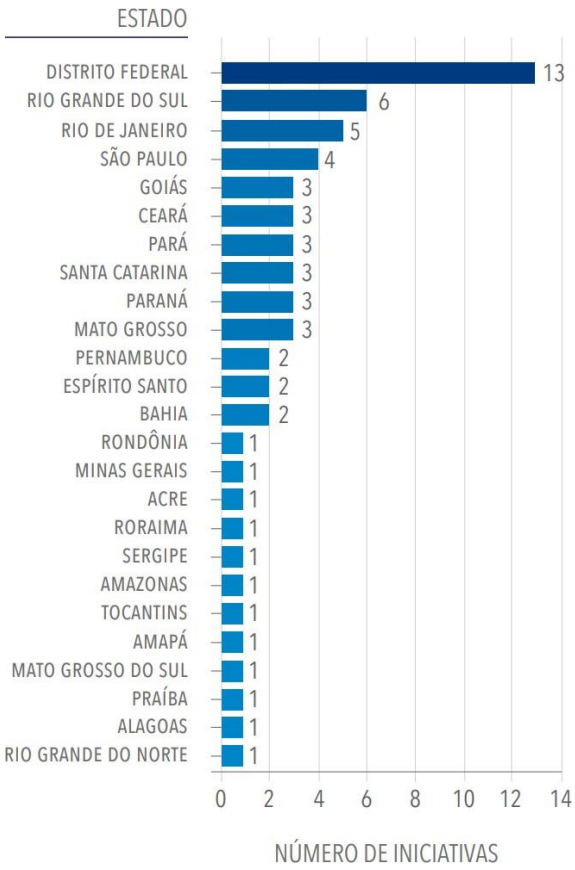
Figura 19 – Iniciativas de uso de Inteligência Artificial por regiões.



Fonte: Salomão (2022).

Na Figura 19 é possível identificar que a região Centro-Oeste tem tendência de uso maior da tecnologia em comparação com as demais regiões do país. Também é possível visualizar que todas as regiões do país há alguma iniciativa de uso de sistemas com Inteligência Artificial. O mesmo pode ser observado em relação aos estados brasileiros, como mostra a Figura 20.

Figura 20 – Iniciativas de uso de Inteligência Artificial por estados.



Fonte: Salomão (2022).

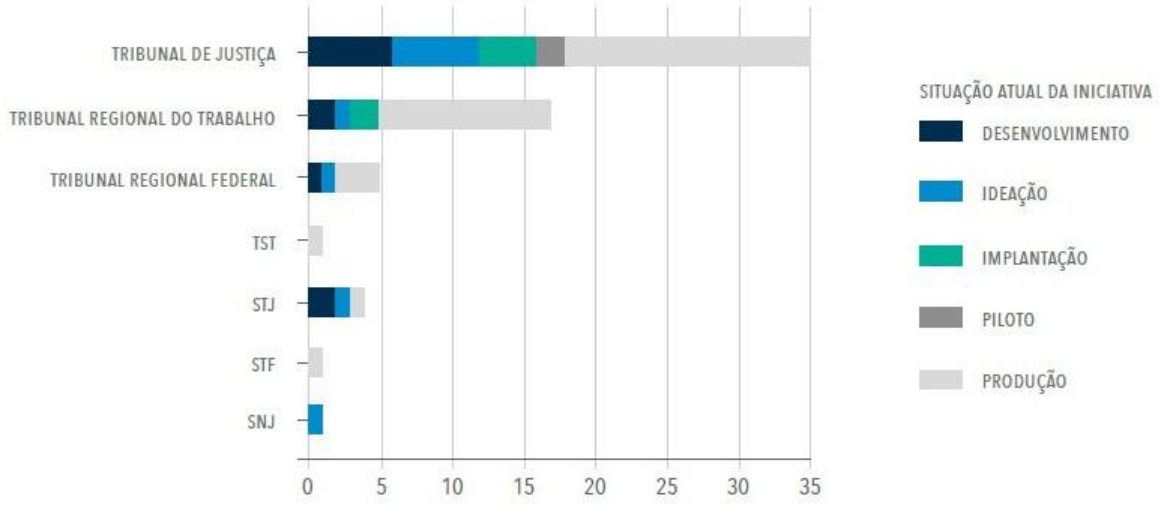
Por sua vez, na Figura 20 é possível identificar as iniciativas de projetos com sistemas inteligente por estados, onde ao menos uma iniciativa existe, a exemplo de estados como Alagoas, Sergipe e Amazonas. É igualmente possível visualizar que dos 26 estados mais o Distrito Federal, apenas 2 estados não adotaram o uso dessa tecnologia, são eles o Piauí e o Maranhão. Desta forma, sugere-se que a tendência de uso da Inteligência Artificial pelo Poder Judiciário constitui-se uma realidade, da qual o Distrito Federal, o Rio Grande de Sul e o Rio de Janeiro são os estados que possuem maior número de iniciativas.

Com a crescente adoção de tecnologias no Poder Judiciário brasileiro, é natural que haja uma preocupação em estabelecer regras para o seu uso. O Conselho Nacional de Justiça emitiu portarias e resoluções para orientar o uso de tecnologias no sistema judicial. No entanto, o ordenamento jurídico brasileiro já observou a necessidade de regulamentação específica para o uso de Inteligência Artificial, o que inclui sua utilização no Judiciário, além das normas jurídicas já aplicáveis. Portanto, é importante que a utilização da Inteligência Artificial seja orientada por princípios éticos e transparentes, para garantir a proteção dos direitos fundamentais e a justiça nas decisões judiciais.

Cada vez mais a preocupação e a cobrança por sistemas transparentes e éticos ganhará destaque em termos de regulação. Todos os instrumentos normativas abarcam tais características

em seu conteúdo de forma explícita ou implícita. Ocorre que as ferramentas de software em desenvolvimento e em produção evidenciam que a tecnologia analisada crescerá no judiciário brasileiro. É que mostra a Figura 21.

Figura 21 – Panorama da Inteligência Artificial no Poder Judiciário brasileiro.



Fonte: Salomão (2022).

Pode-se observar pela Figura 21 que há diversas iniciativas em fases de desenvolvimento, ideação e em teste piloto. Isso indica que diversos sistemas com Inteligência Artificial serão colocados em ambiente de produção de resultados reais. Grande parte desses projetos estão localizados nos tribunais de justiça, o que é corroborado pela Figura 20.

Os indicativos apontam para o aumento real do uso da tecnologia de Inteligência Artificial no judiciário brasileiro. Exatamente por isso que a transparência e a ética desde o seu desenvolvimento se configuram de vital importância. É do próprio interesse dos operadores do direito e da sociedade que exista o controle e a auditabilidade nesses sistemas.

Ainda segundo Salomão et al. (2021), as principais atividades com Inteligência Artificial implementadas nos tribunais brasileiros incluem a extração de informações relevantes de documentos e processos judiciais, a categorização e classificação de documentos para facilitar a organização e recuperação de informações, o agrupamento de processos ou documentos semelhantes para simplificar a análise, e a recuperação aprimorada de informações em grandes volumes de dados judiciais.

Além disso, a Inteligência Artificial é usada na geração automática de documentos legais, como minutas de decisões e sentenças, e também fornece sistemas de recomendação para magistrados, advogados e partes envolvidas, ajudando a encontrar casos, decisões ou jurisprudência relevantes. A Inteligência Artificial também é aplicada no processamento de sinais, como gravações de áudio e vídeo em tribunais, para auxiliar na análise de evidências e provas.

Outras atividades incluem a visualização e exploração avançada de dados judiciais,

otimização de processos judiciais, identificação e indexação automáticas de peças processuais nos autos originários, análises judiciais avançadas com base em dados e inteligência de negócios, além de melhorias na pesquisa jurisprudencial. A Inteligência Artificial também desempenha um papel na triagem automática de documentos jurídicos, na digitalização de acervos físicos, no reconhecimento facial de jurisdicionados e no controle de presença de apenados.

Por fim, a Inteligência Artificial é utilizada na análise de petições iniciais e suas correlações, bem como na prática da jurimetria, que envolve a análise estatística de dados judiciais. Todas essas atividades demonstram como a Inteligência Artificial está transformando o setor jurídico no Brasil, tornando-o mais eficiente, acessível e eficaz para todas as partes envolvidas no processo judicial.

6 CONSIDERAÇÕES FINAIS

A proteção de dados pessoais é uma questão importante, que envolve diversas formas de proteção, incluindo a proteção jurídica. A Lei Geral de Proteção de Dados, no Brasil, trouxe uma proteção jurídica mais ampla em relação ao tratamento de dados pessoais, incluindo a proteção contra tratamento automatizado realizados por sistemas inteligentes. Isso porque a tecnologia computacional tem avançado com o tempo, o que torna viável a implementação de algoritmos complexos que modelam um comportamento inteligente e evolucionário (Inteligência Artificial), permitindo que problemas complexos possam ser resolvidos com mais eficiência e eficácia, tornando possível o processamento de volumes massivos de dados.

No entanto, a automação pode representar um risco significativo aos titulares de dados pessoais, impactando-os diretamente em suas vidas, pois, em muitos casos, existe uma tomada de decisão ou um perfilamento executado somente de forma automatizada sem nenhuma interação humana significativa. Para proteger os titulares dos dados pessoais deste tipo de tratamento automatizado e desumanizado, o legislador incluiu o direito à explicação de decisões automatizadas que impõe ao controlador a obrigação de fornecer informações claras e adequadas sobre o tratamento automatizado.

Este trabalho discute a concretização do direito à explicação na tomada de decisões automatizadas, das quais o entendimento da lógica de processamento resta incompreensível, cenário este aplicado ao processo judicial, levantando preocupações em relação à garantia da observação ao princípio da motivação das decisões judiciais. A transparência no processo de decisão judicial automatizada é fundamental para garantir o direito fundamental ao devido processo legal.

Devido à opacidade no funcionamento dos modelos de aprendizagem de máquinas utilizados até a realização da pesquisa, não é possível fornecer uma explicação completa sobre como o sistema inteligente alcançou determinado resultado em casos de processamento específico, o que resulta em uma concretização parcial do direito à explicação. O titular de dados pessoais teria, neste contexto, informações mínimas para o exercício de seus direitos.

Por essa razão, a adoção de um procedimento para revisão de decisões automatizadas por um ser humano, apesar de também existir riscos, parece o caminho menos danoso para o titular de dados. A decisão resultante da revisão nesses moldes pode ser explicada de forma mais relevante para o titular, uma vez que o ser humano possui habilidade necessária expor seu raciocínio lógico.

Diante disso, o principal objetivo desta pesquisa foi analisar como a Lei Geral de Proteção de Dados pode ser utilizada como via de concretização para o princípio da motivação de decisões judiciais ao lidar com sistemas de Inteligência Artificial através do direito à explicação de decisões automatizadas. Pra isso, foi realizada uma pesquisa baseada em revisão de literatura e em documentos, onde foi possível atingir algumas conclusões.

Dentre as principais conclusões está o uso da transparência e da informação significativa

como fundamentos de uma explicação efetiva e com significado para o titular de dados quanto ao tratamento automatizados que lhe é submetido, do qual a Inteligência Artificial está no escopo. Por causa dos problemas de cunho tecnológicos, que atrapalham uma explicação completa, é apenas possível fornecer parte das informações necessárias para o titular de dados, o impedindo de conhecer o tratamento como um todo. Como consequência, o titular de dados se mantém em um estado de vulnerabilidade constante, impossibilitado de exercer plenamente seus direitos. Além disto, direitos fundamentais, como o privacidade por exemplo, podem ser atingidos e violados.

A explicação, quando fornecida antes da ocorrência do tratamento automatizado e abrangendo todos os aspectos desse processo, incluindo detalhes essenciais como finalidade e base legal, pode ser executada, desde que esteja em conformidade com as disposições legais que visam proteger segredos comerciais e industriais. Isso, portanto, estabelece o nível mínimo de transparência no contexto de proteção de dados e Inteligência Artificial, denominado de transparência parcial.

Além disso, dentro do contexto do processo judicial, quando existir um sistema de Inteligência Artificial no auxílio à tomada de decisão, a explicação significativa vindo do direito à explicação de decisões automatizadas, torna-se basilar para o esclarecimento do como e do porquê determinada proposta de decisão foi criada e assim, quando chancelada pelo juízo, concretizar a motivação dessa decisão judicial. Isso permite que o titular de dados, como parte do processo, possa atacar a decisão adequadamente, o que reflete no devido processo, além do contraditório e da ampla dentre outros direitos e garantias processuais.

Ainda diante do cenário processual, a publicidade da arquitetura do sistema - o que engloba no mínimo os algoritmos e dados utilizados - traz uma confiabilidade para os sistemas de Inteligência Artificial, porque a auditabilidade pode ser realizada, o que se permite a verificação do funcionamento imparcial do sistema na tomada de decisão.

Portanto, por mais desafiador que possa ser a utilização da tecnologia de Inteligência Artificial no tratamento de dados pessoais e especificamente quando aplicada para auxílio ou para a tomada de decisão, não se pode ignorar o avanço destas ferramentas que acabam por ajudar a atividade humana. Contudo, é preciso cautela com o seu uso, pois a tecnologia está longe da perfeição técnica e ainda mais longe de uma incorporação ética em suas regras, cabendo ao ser humano o discernimento para o aproveitamento responsável da tecnologia.

A pesquisa abordou vários aspectos importantes relacionados ao direito à explicação de decisões automatizadas e seu contexto na Lei Geral de Proteção de Dados. Entretanto não foi possível tratar todos os aspectos que são relevantes relacionado ao tema escolhido. Dentre esses aspectos estão: a relação entre o direito à revisão das decisões automatizadas com o direito à explicação tratado neste trabalho; o aprofundamento em relação à responsabilidade civil da utilização de sistemas de Inteligência Artificial dentro do contexto da proteção de dados pessoais; a ampliação do estudo para abranger outros aspectos referentes ao processo judicial com uso da tecnologia de Inteligência Artificial, a exemplo de como esses sistemas lidam com

provas dentro do processo; outro ponto de enfoque será o aprofundamento dos impactos da tecnologia de Inteligência Artificial nos direitos e garantias fundamentais. Isso é essencial para evitar possíveis discriminações resultantes do uso de algoritmos e dados enviesados. Todos os aspectos abordados no parágrafo anterior podem servir como referência para trabalhos futuros.

Este estudo enfrentou várias limitações que moldaram seu escopo e resultados. Em primeiro lugar, o fator tempo impactou a profundidade e refinamento da pesquisa, devido aos requisitos do programa de pós-graduação, especialmente devido aos compromissos com as disciplinas cursadas. Além disso, restrições financeiras limitaram o acesso a alguns trabalhos relacionados, resultando na exclusão de certas referências. Outra limitação importante foi a restrição da área de aplicação da pesquisa, que se concentrou principalmente nas decisões automatizadas no contexto judicial, sem acesso a sistemas específicos com Inteligência Artificial para uma análise mais empírica.

Outras limitações incluíram a concentração da pesquisa no direito à explicação de decisões automatizadas, com menção superficial a direitos relacionados, a exemplo do direito à revisão de decisões automatizadas, à responsabilidade civil por atos realizados por Inteligência Artificial e outros pontos relevantes já citados como possibilidade de pesquisas futuras. Além disso, a pesquisa enfrentou a limitação da imaturidade da literatura em relação ao direito à explicação de decisões automatizadas no contexto da Inteligência Artificial, o que tornou desafiador encontrar soluções satisfatórias aplicáveis ao cenário brasileiro. Todas essas limitações contribuíram para moldar os resultados e as conclusões deste estudo.

REFERÊNCIAS

ADAMSSEN, J. **Inteligência artificial: Como aprendizado de máquina, robótica e automação moldaram nossa sociedade**. Efalon Acies, 2020. ISBN 9788835873785. Disponível em: <https://books.google.com.br/books?id=BN74DwAAQBAJ>.

AFFONSO, Filipe José Medon et al. Inteligência artificial e danos: autonomia, riscos e solidariedade. Universidade do Estado do Rio de Janeiro, 2019.

AGGARWAL, Charu. **Neural Networks and Deep Learning: A Textbook**. [S.l.]: Springer International Publishing, 2018.

ALAN, Poole David Mackworth. Gr: Computational intelligence: a logical approach.(1998). **Google scholar google scholar digital library digital library**, 1998.

ALBUQUERQUE, Paula Falcão; JÚNIOR, Marcos Ehrhardt. Quais as consequências do não atendimento do recall pelo consumidor? **Revista Jurídica Luso-Brasileira**, n. 3, 2018.

ALPAYDIN, Ethem. **Introduction to machine learning**. 4rd. ed. [S.l.]: MIT press, 2020.

ANPD, Autoridade Nacional de Proteção de Dados. **Análise preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial**. 2023. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/analise-preliminar-do-pl-2338_2023-formatado-ascom.pdf>.

Autoridade Nacional de Proteção de Dados (ANPD). **Guia Orientativo para Definições dos Agentes de Tratamento de Dados Pessoais e do Encarregado, Versão 2.0**. [S.l.], 2022. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/guias-e-manuais/guias-e-manuais-1/guias-e-manuais/guia-definicoes-dos-agentes-de-tratamento-e-do-encarregado.pdf/view>.

BARBIERI, Carlos. **Governança de Dados: Práticas, Conceitos e Novos Caminhos**. Rio de Janeiro: Alta Books, 2019. 288 p.

BARBOSA, Tales Schmidke; TONIAZZO, Daniela Wendt; RUARO, Regina Linden. O direito à explicação nas decisões automatizadas: uma abordagem comparativa entre o ordenamento brasileiro e europeu. **Revista internacional CONSINTER de direito**, p. 55–69, 2021.

BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. Fairness in machine learning. **Nips tutorial**, v. 1, p. 2017, 2017.

BARRETO, Angela Maria. Informação e conhecimento na era digital. **Transinformação**, Campinas, v. 17, n. 2, p. 111–122, 2005.

BARROSO, Luís Roberto. Neoconstitucionalismo e constitucionalização do direito: o triunfo tardio do direito constitucional no brasil. **Bol. Fac. Direito U. Coimbra**, HeinOnline, v. 81, p. 233, 2005.

BERNARDI, Renato; PIEROBON, Flávio. A constituição do estado e da sociedade: uma análise da força normativa da constituição brasileira de 1988. **Revista do Direito Público, Brasil, Londrina**, v. 9, n. 1, p. 55–72, 2014.

BIONI, Bruno; DIAS, Daniel. Responsabilidade civil na proteção de dados pessoais: construindo pontes entre a lei geral de proteção de dados pessoais e o código de defesa do consumidor. **Civilistica. com**, v. 9, n. 3, p. 1–23, 2020.

BIONI, Bruno Ricardo; LUCIANO, Maria. O PRINCÍPIO DA PRECAUÇÃO NA REGULAÇÃO DE INTELIGÊNCIA ARTIFICIAL: SERIAM AS LEIS DE PROTEÇÃO DE DADOS O SEU PORTAL DE ENTRADA? In: **Inteligência Artificial e Direito**. [S.l.]: Revista dos Tribunais, 2019. p. 720. ISBN 8553217299.

BLACKLAWS, Christina. Algorithms: transparency and accountability. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 376, n. 2128, p. 20170351, sep 2018. ISSN 1364-503X. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rsta.2017.0351>.

BRACHMAN, Ronald J.; LEVESQUE, Hector J. **Knowledge Representation and Reasoning**. [S.l.]: Elsevier, 2004.

BRASIL. **Código de Processo Penal**. Presidência da República, 1941. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto-lei/Del3689.htm>.

____. **Lei Complementar nº 35, de 14 de março de 1979**. 1979. <http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp35.htm>.

____. **Constituição da República Federativa do Brasil**. Brasília: Senado Federal, 1988.

____. **Lei nº 9.609, de 19 de fevereiro de 1998**. 1998. <http://www.planalto.gov.br/ccivil_03/Leis/L9609.htm>. Acesso em: 11 abr. 2023.

____. **LEI No 10.406, DE 10 DE JANEIRO DE 2002. Código Civil.** Brasília: Senado Federal, 2002. Disponível em: <http://www.planalto.gov.br/CCivil_03/Leis/2002/L10406.htm>.

____. **LEI Nº 12.965, DE 23 DE ABRIL DE 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil.** Brasília: [s.n.], 2014. Online p. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm>.

____. **LEI Nº 13.105, DE 16 DE MARÇO DE 2015. Código de Processo Civil**. Brasília: [s.n.], 2015.

____. **Lei 13709 de 2018, Lei Geral de Proteção de Dados**. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>.

____. **Projeto de Lei nº 5051, de 2019. Estabelece os princípios para o uso da Inteligência Artificial no Brasil**. 2019a. <https://www25.senado.leg.br/web/atividade/materias/-/materia/137689>. Acesso em: 11 de abril de 2023.

____. **Projeto de Lei nº 5691 de 2019. Institui a Política Nacional de Inteligência Artificial**. 2019b.

____. **Projeto de Lei nº 4496 de 2019. Altera a Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais – LGPD), para definir a expressão "decisão automatizada".** 2019c.

____. **Substitutivo do Projeto de Lei 21/2020**. Brasília: Senado Federal, 2020. 11 p.

____. **Projeto de Lei nº 872 de 2021. Dispõe sobre o uso da Inteligência Artificial**. 2021.

____. **Senado Federal. Projeto de Lei 2338/2023, de 03 de maio de 2023. Dispõe sobre o uso da Inteligência Artificial**. 2023. Acesso em: 11 mai. 2023. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9347593&ts=1683827933930&disposition=inline&_gl=1*eriul*_ga*MjUxMTY1ODU4LjE2ODM4Mjk5Mjc.*_ga_CW3ZH25XMK*MTY4NDI0ODg3OC40LjAuMTY4NDI0ODg3OC4wLjAuMA>.

BRENNAN, Tim; DIETERICH, William. **Correctional Offender Management Profiles for Alternative Sanctions (COMPAS)**. John Wiley & Sons, Ltd, 2018. 49-75 p. ISBN 9781119184256. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119184256.ch3>.

BUTTERWORTH, Michael. The ICO and artificial intelligence: The role of fairness in the GDPR framework. **Computer Law & Security Review**, Elsevier Ltd, v. 34, n. 2, p. 257–268, apr 2018. ISSN 02673649. Disponível em: <https://doi.org/10.1016/j.clsr.2018.01.004https://linkinghub.elsevier.com/retrieve/pii/S026736491830044X>.

CELESTE, Edoardo; SANTARÉM, Paulo Rená da Silva. Constitucionalismo digital: mapeando a resposta constitucional aos desafios da tecnologia digital. **Revista Brasileira de Direitos Fundamentais & Justiça**, v. 15, n. 45, p. 63–91, 2021.

CETAX. **Deep Learning: O que é, conceitos e definições**. 2022. Disponível em: <https://www.cetax.com.br/blog/o-que-e-deep-learning/>.

CHEN, Min et al. Data mining for the internet of things: Literature review and challenges. **International Journal of Data Mining and Knowledge Management Process**, v. 4, n. 3, p. 1–17, 2014.

CIELO, PFLD et al. Uma leitura dos princípios da prevenção e da precaução e seus reflexos no direito ambiental. **Revista CEPPG [internet]**, v. 26, n. 1, p. 196–207, 2012.

CJF, Conselho da Justiça Federal. **I Jornada de Direito Administrativo Enunciados Aprovados**. 2020. 1-23 p. Disponível em: <https://dizerodireitodotnet.files.wordpress.com/2020/08/enunciados-aprovados-i-jda-vf.pdf>.

CNJ. **Portaria nº. 271. Regulamenta o uso de Inteligência Artificial no âmbito do Poder Judiciário.** Brasília: Conselho Nacional de Justiça - CNJ, 2020. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3613>.

____. **Resolução nº. 332, de 21 de Agosto de 2020**. Brasília: Conselho Nacional de Justiça -CNJ, 2020. 1–11 p. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>.

CNJ, Conselho Nacional de Justiça. **Código de Ética da Magistratura Nacional**. 2008. <http://www.cnj.jus.br/images/resolucoes/codigo_etica.pdf>.

COPPIN, Ben. Artificial intelligence illuminated / b. coppin. p. 740, 01 2004.

CROCKETT, Keeley; GOLTZ, Sean; GARRATT, Matt. GDPR Impact on Computational Intelligence Research. In: **2018 International Joint Conference on Neural Networks (IJCNN)**. IEEE, 2018. v. 2018-July, p. 1–7. ISBN 978-1-5090-6014-6. Disponível em: <https://ieeexplore.ieee.org/document/8489614/>.

CRUZ, Nuno Ferreira da et al. Measuring local government transparency. **Public Management Review**, Taylor & Francis, v. 18, n. 6, p. 866–893, 2016.

DERMEVAL, Diego et al. A Systematic Review on the Use of Ontologies in Requirements Engineering. In: **2014 Brazilian Symposium on Software Engineering**. IEEE, 2014. p. 1–10. ISBN 978-1-4799-4223-7. ISSN 0975- 5861. Disponível em: <http://ieeexplore.ieee.org/document/6943477/>.

Didier Jr, Fredie; BRAGA, Paula Sarno; OLIVEIRA, Rafael Alexandria de. Curso de direito processual civil: teoria da prova, direito probatório, ações probatórias, decisão, precedente, coisa julgada e antecipação dos efeitos da tutela. **Salvador: Jus Podivm**, v. 2, 2015.

DING, Wei et al. Knowledge-based approaches in software documentation: A systematic literature review. **Information and Software Technology**, Elsevier B.V., v. 56, n. 6, p. 545–567, jun 2014. ISSN 09505849. Disponível em: <http://dx.doi.org/10.1016/j.infsof.2014.01.008http://linkinghub.elsevier.com/retrieve/pii/S0950584914000196>.

DIÓGENES JÚNIOR, José Eliaci Nogueira. Gerações ou dimensões dos direitos fundamentais. **Âmbito Jurídico, Rio Grande, XV**, n. 100, p. 571–572, 2012.

DOSHI-VELEZ, Finale et al. Accountability of AI Under the Law: The Role of Explanation. **SSRN Electronic Journal**, p. 1–15, 2017. ISSN 1556-5068. Disponível em: <https://www.ssrn.com/abstract=3064761>.

DYBÅ, Tore; DINGSØYR, Torgeir. Empirical studies of agile software development: A systematic review. **Information and Software Technology**, v. 50, n. 9-10, p. 833–859, aug 2008. ISSN 09505849. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0950584908000256>.

ECO, Umberto et al. **Interpretação e superinterpretação**. [S.l.]: Martins Fontes São Paulo, 1993.

EDWARDS, Lilian; VEALE, Michael. Slave to the Algorithm? Why a Right to Explanationn is Probably Not the Remedy You are Looking for. **SSRN Electronic Journal**, v. 2017, 2017. ISSN 1556-5068. Disponível em: <https://www.ssrn.com/abstract=2972855>.

____. Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"? **IEEE Security & Privacy**, v. 16, n. 3, p. 46–54, may 2018. ISSN 1540-7993. Disponível em: <https://ieeexplore.ieee.org/document/8395080/>.

EU, European Union. **General Data Protection Regulation (GDPR) – Final text neatly arranged**. 2016. Disponível em: <https://gdpr-info.eu/>.

FELZMANN, Heike et al. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. **Big Data & Society**, v. 6, n. 1, p. 205395171986054, jan 2019. ISSN 2053-9517. Disponível em: <http://journals.sagepub.com/doi/10.1177/2053951719860542>.

FERRARI, Isabela. Accountability de algoritmos: a falácia do acesso ao código e caminhos para uma explicabilidade efetiva. **Inteligência Artificial: 3º Grupo de Pesquisa do ITS, ITS-Instituto de Tecnologia e Sociedade do Rio**, 2018.

GANEK, A. G.; CORBI, T. A. The dawning of the autonomic computing era. **IBM Systems Journal**, v. 42, n. 1, p. 5–18, 2003. ISSN 0018-8670. Disponível em: <http://ieeexplore.ieee.org/document/5386835/>.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GILLET, Sérgio Augusto da Costa; PORTELA, Vinícius José Rockenbach. Breves conexões entre a motivação das decisões judiciais e o campo da inteligência artificial. **Cadernos de Direito**, v. 18, n. 34, p. 153–171, 2018.

GONDIM, Glenda Gonçalves. A responsabilidade civil no uso indevido dos dados pessoais. **Revista IBERC**, v. 4, n. 1, p. 19–34, 2021.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento de Imagens Digitais**. 3rd. ed. [S.l.]: Edgard Blücher, 2008.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016.

GOODMAN, Bryce. A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection. **29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.**, n. Nips, p. 1–7, 2016.

GOODMAN, Bryce; FLAXMAN, Seth. European union regulations on algorithmic decision making and a "right to explanation". **AI Magazine**, v. 38, n. 3, p. 50–57, 2017. ISSN 07384602.

GUARDA, Paolo. "Ok Google, am I sick?": artificial intelligence, e-health, and data protection regulation. **BioLaw Journal – Rivista di BioDiritto**, v. 1, p. 359–375, 2019. ISSN 2284-4503.

GUEDES, Gisela Sampaio da Cruz. Regime de responsabilidade adotado pela lei de proteção de dados brasileira. **Caderno especial LGPD**, p. 167–182, 2019.

HENNESSY, John L; PATTERSON, David A. **Organização e Projeto de Computadores: a interface hardware/software**. [S.l.]: Elsevier Brasil, 2014. v. 4.

ISO, ABNT-NBR. Iec 27032: 2012: Tecnologia da informação-técnicas de segurança-diretrizes para segurança cibernética. **Rio de Janeiro: ABNT**, 2015.

JURAFSKY, Daniel; MARTIN, James H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 3rd. ed. [S.l.]: Pearson, 2019.

KAMARINOU, Dimitra; MILLARD, Christopher; SINGH, Jatinder. Machine Learning with Personal Data: Profiling, Decisions and the EU General Data Protection Regulation. **Queen Mary School of Law Legal Studies Research Paper No. 247/2016**, p. 1–7, 2016. Disponível em: <http://www.mlandthelaw.org/papers/kamarinou.pdf>.

KAMINSKI, Margot E. The Right to Explanation, Explained. **SSRN Electronic Journal**, v. 34, n. 1, 2018. ISSN 1556-5068. Disponível em: <https://www.ssrn.com/abstract=3196985>.

KIM, Pauline T. Auditing Algorithms for Discrimination. **University of Pennsylvania Law Review**, v. 166, n. 1, 2017. ISSN 0041-9907. Disponível em: <https://scholarship.law.upenn.edu/penn_law_review_online/vol166/iss1/10>.

KIM, Tae Wan; ROUTLEDGE, Bryan R. Informational Privacy, A Right to Explanation, and Interpretable AI. In: **2018 IEEE Symposium on Privacy-Aware Computing (PAC)**. IEEE, 2018. p. 64–74. ISBN 978-1-5386-8442-9. Disponível em: <https://ieeexplore.ieee.org/document/8511831/>.

KINGSTON, John. Using artificial intelligence to support compliance with the general data protection regulation. **Artificial Intelligence and Law**, Springer Netherlands, v. 25, n. 4, p. 429–443, 2017. ISSN 15728382.

KITCHENHAM, B. **Procedures for performing systematic reviews**. [S.l.], 2004. v. 3, n. 3, 33 p.

KITCHENHAM, B; CHARTERS, S. **Guidelines for Performing Systematic Literature Reviews in Software Engineering**. [S.l.], 2007.

LAGE, Fernanda de Carvalho. Manual de inteligência artificial no direito brasileiro. **Salvador: JusPodivm**, 2021.

LAROSE, D. T.; LAROSE, C. D. **Discovering knowledge in data: an introduction to data mining**. 2nd. ed. [S.l.]: John Wiley & Sons, 2014.

LARSON, Ron; FARBER, Betsy. **Elementary Statistics: Picturing the World**. 7th. ed. Boston: Pearson, 2019.

LARSSON, Stefan; HEINTZ, Fredrik. Transparency in artificial intelligence. **Internet Policy Review**, v. 9, n. 2, 2020.

LAVALLE, Steven M. **Planning Algorithms**. Cambridge: Cambridge University Press, 2006.

LEONARDO, César Augusto Luiz; ESTEVÃO, Roberto da Freiria. Inteligência artificial, motivação das decisões, hermenêutica e interpretação: alguns questionamentos a respeito da inteligência artificial aplicada ao direito. **Revista Em Tempo**, v. 20, n. 1, 2020.

LIMA FILHO, Maxwell Morais de. O experimento de pensamento do quarto chinês: a crítica de john searle à inteligência artificial forte. Argumentos Revistas de Filosofia, 2010.

LIMA, Iara Menezes; LANÇA, João André Alves. A força normativa da constituição e os limites à mutação constitucional em konrad hesse. **Revista da Faculdade de Direito da UFMG**, n. 62, p. 275–304, 2013.

LOPES, Giovana F. P. O "DIREITO À EXPLICAÇÃO" DE DECISÕES AUTOMATIZADAS NO ÂMBITO DO GDPR. In: **Data Protection Law in the EU: Roles, Responsibilities and Liability**. Intersentia, 2019. p. 279–324. Disponível em: <https://www.cambridge.org/core/product/identifier/CBO9781780688459A070/type/book_part>.

LUCCA, Rodrigo Ramina de. O dever de motivação das decisões judiciais. **Salvador: Juspodivm**, 2015.

LUGER, George F. **Inteligência Artificial. Tradução de Daniel Vieira**. [S.l.]: São Paulo: Pearson Education do Brasil, 2013.

LYNSKEY, Orla. Criminal justice profiling and EU data protection law: precarious protection from predictive policing. **International Journal of Law in Context**, v. 15, n. 2, p. 162–176, jun 2019. ISSN 1744-5523. Disponível em: <https://www.cambridge.org/core/product/identifier/S1744552319000090/type/journal_article>.

LYRA, Carlos Eduardo de Sousa; MOGRABI, Gabriel José Corrêa; EL-HANI, Charbel Niño. O naturalismo biológico de searle e a relação mente-cérebro. **Psicologia: Teoria e Pesquisa**, SciELO Brasil, v. 32, p. 7–15, 2016.

MAHDAVI-HEZAVEHI, Sara; GALSTER, Matthias; AVGERIOU, Paris. Variability in quality attributes of service-based software systems: A systematic literature review. **Information and Software Technology**, Elsevier B.V., v. 55, n. 2, p. 320–343, feb 2013. ISSN 09505849. Disponível em: <http://dx.doi.org/10.1016/j.infsof.2012.08.010http://linkinghub.elsevier.com/retrieve/pii/S0950584912001772>.

MALGIERI, Gianclaudio. Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations. **Computer Law & Security Review**, Elsevier Ltd, v. 35, n. 5, p. 105327, oct 2019. ISSN 02673649. Disponível em: <https://doi.org/10.1016/j.clsr.2019.05.002https://linkinghub.elsevier.com/retrieve/pii/S0267364918303753>.

MALGIERI, Gianclaudio; COMANDÉ, Giovanni. Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. **International Data Privacy Law**, v. 7, n. 4, p. 243–265, nov 2017. ISSN 2044-3994. Disponível em: <http://academic.oup.com/idpl/article/7/4/243/4626991>.

MEDON, Filipe. Inteligência artificial e responsabilidade civil: autonomia, riscos e solidariedade. **Salvador: Editora JusPodivm**, 2022.

MENDONÇA, Fernanda Graebin. O DIREITO À AUTODETERMINAÇÃO INFORMATIVA: A (DES)NECESSIDADE DE CRIAÇÃO DE UM NOVO DIREITO FUNDAMENTAL PARA A PROTEÇÃO DE DADOS PESSOAIS NO BRASIL. In: **XI Seminário Internacional de Demandas Sociais e Políticas Públicas na Sociedade Conteporânea**. [S.l.: s.n.], 2014.

MITIDIERO, Daniel Francisco. O ônus da prova e seus inimigos. **Revista de Processo**, v. 306, p. 17–47, 2020.

MONTEIRO, Cláudia Servilha. Fundamentos para uma teoria da decisão judicial. p. 6104–6125, 2007. Disponível em: <http://www.publicadireito.com.br/conpedi/manaus/arquivos/anais/bh/claudia_servilha_monteiro.pdf>.

MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados do Brasil? **Instituto Igarapé**, v. 39, p. 1–27, 2018. Disponível em: <https://igarape.org.br/wp-content/uploads/2018/12/Existe-um-direito-a-explicacao-na-Lei-Geral-de-Protecao-de-Dados-no-Brasil.pdf>.

MOREIRA, Davi Antônio Gouvêa Costa. A relação entre constituição e processo no estado contemporâneo: Constitucionalização do direito processual e neoprocessualismo. **Olhares Plurais**, v. 2, n. 11, p. 26–38, 2014.

MOREIRA, S. V. **Análise documental como método e como técnica. In: Jorge Duarte; Antônio Barros. (Org). Métodos e técnicas de pesquisa em comunicação**. São Paulo: Atlas, 2005.

MOUGIAKOU, Eirini; PAPADIMITRIOU, Spyros; VIRVOU, Maria. Automated Decision Making and Personal Data Protection in Intelligent Tutoring Systems: Design Guidelines. In: VIRVOU, Maria; KUMENO, Fumihiro; OIKONOMOU, Konstantinos (Ed.). **Springer Nature Switzerland**. Cham: Springer International Publishing, 2019, (Smart Innovation, Systems and Technologies, v. 108). p. 231–241. ISBN 978-3-319-97678-5. Disponível em: <http://link.springer.com/10.1007/978-3-319-97679-2_24>.

MULHOLLAND, Caitlin. Responsabilidade civil por danos causados pela violação de dados sensíveis e a lei geral de proteção de dados pessoais.(lei 13.709/2018). **MARTINS, Guilherme Magalhães; ROSENVALD, Nelson (Coords.). Responsabilidade civil e novas tecnologias. Indaiatuba, SP. Editora Foco**, 2020.

NOGUEIRA, Pedro Henrique Pedrosa. Os limites lingüístico-legislativos da discricionariedade judicial. **Revista de Informação Legislativa**, v. 181, p. 313–325, 2009.

NORVIG, Peter; RUSSELL, Stuart. Inteligência artificial. **Tradução: Regina Célia Simille de Macedo. Consultoria Editorial e Revisão técnica: Dr. Flávio Soares Corrêa da Silva, Dra. Leliane Nunes de Barros and Dra. Renata Wassermann**, v. 3, p. 13–31, 2013.

NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. v. 285, n. 2018, p. 421–447, 2018.

NUNES, Vanessa; CAPPELLI, Claudia; RALHA, Célia G. Transparency in information systems. **Sociedade Brasileira de Computação**, 2017.

OLIVEIRA, Leonardo Alves de. A sétima dimensão dos direitos fundamentais. **6Novembro 2016**, p. 395, 2016.

Organização das Nações Unidas, ONU. **Pacto Internacional dos Direitos Civis e Políticos**. ONU, 1966. Disponível em: <https://www2.camara.leg.br/legin/fed/lei/1970-1979/lei-6929-81-27-outubro-1981-383442-norma-pl.html>.

PATTERSON, David A; HENNESSY, John L. **Computer organization and design ARM edition: the hardware software interface**. [S.l.]: Morgan kaufmann, 2016.

PEIXOTO, Fabiano Hartmann; COUTINHO, Marina de Alencar Araripe. Inteligência artificial e regulação. **Revista Em Tempo**, v. 19, n. 1, 2020.

PEREIRA, Luciano Meneguetti. As dimensões de direitos fundamentais e necessidade de sua permanente reconstrução enquanto patrimônio de todas as gerações. **Revista Conexão Eletrônica, Três Lagoas**, v. 10, n. 1, p. 779–803, 2013.

PESAPANE, Filippo et al. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. **Insights into Imaging**, v. 9, n. 5, p. 745–753, oct 2018. ISSN 1869-4101. Disponível em: <https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0645-y>.

POLITOU, Eugenia; ALEPIS, Efthimios; PATSAKIS, Constantinos. Profiling tax and financial behaviour with big data under the GDPR. **Computer Law & Security Review**, Elsevier Ltd, v. 35, n. 3, p. 306–329, may 2019. ISSN 02673649. Disponível em: <https://doi.org/10.1016/j.clsr.2019.01.003https://linkinghub.elsevier.com/retrieve/pii/S026736491830133X>.

QUINLAN, J Ross. Induction of decision trees. **Machine learning**, Springer, v. 1, n. 1, p. 81–106, 1986.

RAMAKRISHNAN, Raghu; GEHRKE, Johannes. **Sistemas de gerenciamento de banco de dados-3**. [S.l.]: AMGH Editora, 2008.

RAUBER, Thomas Walter. Redes neurais artificiais. **Universidade Federal do Espírito Santo**, v. 29, 2005.

RIBEIRO, Darci; MAZZOLA, Marcelo. Processo e novas tecnologias nos tribunais: Desafios e perspectivas. **JOTA Info**, 2019. Acesso em 21 de julho de 2023. Disponível em: <https: //amaerj.org.br/wp-content/uploads/2019/11/Processo-e-novas-tecnologias-nos-tribunais_ -desafios-e-perspectivas-JOTA-Info.pdf>.

RIBEIRO, Sergio. **EXTRAÇÃO DE CARACTERÍSTICAS DE IMAGENS APLICADA À DETECÇÃO DE GRÃOS ARDIDOS DE MILHO**. Tese (Doutorado), 02 2015.

RODRIGUES, Andressa Conterno. As dimensões dos direitos fundamentais e sua eficácia nas relações interprivadas. **Revista Direito & Inovação**, v. 1, n. 1, p. 62–74, 2013.

ROQUE, Andre Vasconcelos; SANTOS, Lucas Braz Rodrigues dos. INTELIGÊNCIA ARTIFICIAL NA TOMADA DE DECISÕES JUDICIAIS: TRÊS PREMISSAS BÁSICAS. **Revista Eletrônica de Direito Processual – REDP**, Rio de Janeiro, v. 22, p. 58–78, 2021.

SALOMÃO, Luis Felipe et al. Inteligência artificial: tecnologia aplicada à gestão dos conflitos no âmbito do poder judiciário brasileiro. **Centro de Inovação e Administração e Pesquisa do Judiciário. Rio de Janeiro: FGV Conhecimento**, 2021.

SALOMÃO, Luis Felipe. **Inteligência Artificial: Tecnologia Aplicada À Gestão Dos Conflitos No Âmbito Do Poder Judiciário Brasileiro - 2 Fase**. Carolrhoda Books, 2022. 1-266 p. Disponível em: <https://ciapj.fgv.br/sites/ciapj.fgv.br/files/relatorio_ia_2fase.pdf>.

SANTOLIM, Cesar. Nexo de causalidade e prevenção na responsabilidade civil. **Revista da AJURIS**, v. 41, n. 136, 2014.

SANTOS, Diego Ferreira dos. A constitucionalização do direito civil. **Revista Iurisprudentia**, v. 10, n. 19, 2021.

SANTOS, Danilo Tenório dos; PINTO, Robson de Oliveira; ADAME, Alcione Orientadora. Acesso. à água como direito fundamental e a limitação que seu acesso impõe ao direito de propriedade. 2016.

SANTOS, Leonardo Fernandes dos. Quarta geração/dimensão dos direitos fundamentais: pluralismo, democracia e o direito de ser diferente. IOB; IDP, 2010.

SARLET, Ingo Wolfgang. Proteção de dados pessoais como direito fundamental na constituição federal brasileira de 1988. **Direitos Fundamentais & Justiça**, 2020.

SELBST, Andrew D.; BAROCAS, Solon. The Intuitive Appeal of Explainable Machines. **SSRN Electronic Journal**, v. 87, p. 1085–1139, 2018. ISSN 1556-5068. Disponível em: <https://www.ssrn.com/abstract=3126971>.

SELBST, Andrew D.; POWLES, Julia. Meaningful information and the right to explanation. **International Data Privacy Law**, v. 7, n. 4, p. 543–556, dec 2017. ISSN 2210-5433. Disponível em: <http://link.springer.com/10.1007/s13347-017-0263-5>.

SEMIDÃO, Rafael Aparecido Moron. **DADOS, INFORMAÇÃO E CONHECIMENTO ENQUANTO ELEMENTOS DE COMPREENSÃO DO UNIVERSO CONCEITUAL DA CIÊNCIA DA INFORMAÇÃO: CONTRIBUIÇÕES TEÓRICAS**. 198 p. Tese (Dissertação) — Universidade Estadual Paulista, São Paulo, 2014.

SENADO FEDERAL. **Proposta de Emenda À Constinuição nº 17 de 2019**. Brasília: [s.n.], 2019.

SETZER, Valdemar W. Dado, informação, conhecimento e competência. **DataGramaZero Revista de Ciência da Informação, n. 0**, v. 28, 1999.

SILVA, Virgilio Afonso Da. Direitos fundamentais. **Conteúdo essencial, restrições e**, 2009.

SINGH, Jatinder; COBBE, Jennifer; NORVAL, Chris. Decision Provenance: Harnessing Data Flow for Accountable Systems. **IEEE Access**, v. 7, p. 6562–6574, 2019. ISSN 2169-3536. Disponível em: <https://ieeexplore.ieee.org/document/8579125/>.

SIRIHAL, Adriana Bogliolo; LOURENÇO, Cíntia de Azevedo. Informação e conhecimento: aspectos filosóficos e informacionais. **Informação & Sociedade**, Universidade Federal da Paraíba-Programa de Pós-Graduação em Ciência da …, v. 12, n. 1, 2002.

SOMBRA, Thiago Luis Santos. Fundamentos da regulação da privacidade e proteção de dados pessoais. **São Paulo: Thomson Reuters Brasil**, 2019.

STALLINGS, William. **Arquitetura e organização de computadores**. 8. ed. São Paulo: Pearson Prentice Hall, 2010.

STF, Supremo Tribunal Federal. **STF lança RAFA, ferramenta de Inteligência Artificial para classificar ações na Agenda 2030 da ONU**. 2022. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=486889&ori=1>.

____. **STF finaliza testes de nova ferramenta de Inteligência Artificial**. 2023. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=507120&ori=1>.

SUTTON, Richard S; BARTO, Andrew G. **Reinforcement Learning: An Introduction**. [S.l.]: MIT press, 2018.

TAMBOSI, Paulo Vitor Petris. **Responsabilidade civil pelo tratamento de dados pessoais conforme a Lei Geral de Proteção de Dados (LGPD): subjetiva ou objetiva?** 112 p. Tese (Monografia) — UNIVERSIDADE FEDERAL DE SANTA CATARINA, Florianópolis, 2021.

UNIÃO EUROPEIA. **Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data**. Brussels, Belgium: Official Journal of the European Union, 1995.

____. **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)**. 2016. Disponível em: <https: //eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679>.

VALE, Luís Manoel Borges do. A tomada de decisão por máquinas: a proibição, no direito, de utilização de algoritmos não supervisionados. **Revista de Direito Bancário e do Mercado de Capitais (RBDM)**, 2019.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Transparent, explainable, and accountable ai for robotics. **Science robotics**, American Association for the Advancement of Science, v. 2, n. 6, p. eaan6080, 2017.

____. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. **International Data Privacy Law**, v. 7, n. 2, p. 76–99, may 2017. ISSN 2044-3994. Disponível em: <https://academic.oup.com/idpl/article-lookup/doi/10.1093/idpl/ipx005>.

WACHTER, Sandra; MITTELSTADT, Brent; RUSSELL, Chris. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. **SSRN Electronic Journal**, p. 1–52, 2017. ISSN 1556-5068. Disponível em: <https://www.ssrn.com/abstract=3063289>.

WARAT, Luis Alberto. **O Direito e Sua Linguagem**. [S.l.]: Saraiva, 1995.

ZANATTA, Rafael A. F. Perfilização, Discriminação e Direitos: do Código de Defesa do Consumidor à Lei Geral de Proteção de Dados Pessoais. n. February, 2019. Disponível em: <http://arxiv.org/abs/1606.08813\%0Ahttp://dx.doi.org/10.1609/aimag.v38i3.2741>.

ZHU, Xiaojin; GOLDBERG, Andrew B; ZHU, Xinhua. **Introduction to semi-supervised learning**. Morgan & Claypool, 2009. v. 3. (Synthesis Lectures on Artificial Intelligence and Machine Learning, 1). Disponível em: <https://doi.org/10.2200/S00196ED1V01Y200906AIM006>.

APÊNDICE A – DIREITO À EXPLICAÇÃO DE DECISÕES AUTOMATIZADAS POR INTELIGÊNCIA ARTIFICIAL

A revisão de literatura realizada teve dois propósitos: 1) a identificação, na literatura, da viabilidade do problema de pesquisa selecionado para enfrentamento neste projeto de pesquisa e; 2) o entendimento do estado da arte relativo ao objeto do trabalho de pesquisa. Para isto, foi realizada uma Revisão Sistemática de Literatura (RLS).

A.1 Revisão de literatura - Direito à explicação em decisões automatizadas realizadas por Inteligência Artificial

Uma Revisão Sistemática de Literatura é um meio de se identificar, avaliar e interpretar resultados de pesquisas disponíveis que possuem relação com determinadas questões de pesquisas, fenômenos ou áreas temáticas. Uma RSL, portanto, tem como objetivo principal para sua condição a coleta de evidências que podem levar a algumas conclusões KITCHENHAM; CHARTERS (2007). Isto permite, muitas vezes, identificar o estado da arte de um determinado tema. De acordo com o trabalho citado, uma RSL é composta por três fases distintas e complementares, a saber: I - Na fase de Planejamento, a qual são definidas as questões de pesquisa, é desenvolvido um protocolo de revisão; II - Na fase de Condução da Revisão são identificados, selecionados e avaliados os trabalhos de maior relevância, em conformidade com o protocolo validado na fase anterior, para posterior extração os dados necessários, os quais serão sistematizados e; III - Na fase de Documentação da Revisão é gerado o relatório da revisão e validação.

A.2 Protocolo de pesquisa

Na fase de planejamento, um protocolo de pesquisa foi desenvolvido com base nas diretrizes propostas por KITCHENHAM (2004), e alimentados no software StArt, tendo como objetivo geral obter o atual estado da arte em relação ao tema, criando um resumo sobre decisões automatizadas e perfilamento por softwares complexos e inteligentes (Inteligência Artificial).

A.2.1 Questões de pesquisa

Com o fim de guiar esta RSL, uma questão principal de pesquisa (QP) foi estabelecida e duas questões secundárias foram definidas e são apresentadas a seguir, listando-as com suas respectivas descrições e motivações.

Tabela 3 – Questões de pesquisa.

Questões de Pesquisa	**Descrição e Motivação**
QP1: Como os pesquisadores estão abordando o direito à explicação de decisões automatizadas realizados por algoritmos de Inteligência Artificial?	Essa pergunta direciona toda a RSL e dá um ponto de início para o corte metodológico na pesquisa. Saber como os pesquisadores estão abordando o direito à explicação de processamentos automatizados por Inteligência Artificial pode indicar se a área de estudo está de alguma forma amadurecida ou em processo de amadurecimento.
QP2: Quais problemas enfrentados relacionados ao direito à explicação em decisões automatizadas?	Esta pergunta tem como objetivo principal identificar os principais problemas relatados na literatura sobre o tema da revisão, o que pode permitir observar lacunas existentes.
QP3: Quais as soluções propostas para o enfrentamento dos problemas levantados?	Esta pergunta tem como finalidade levantar as soluções propostas para enfrentamento dos problemas levantados na QP2.

Fonte: Autor.

As questões de pesquisa levantadas no Quadro 1 partem do problema de pesquisa definido neste projeto. Desta forma, guiarão as demais etapas da revisão, auxiliando no apontamento das palavras-chave para as buscas até a sistematização dos resultados.

A.2.2 Fontes de pesquisa e *Strings* de busca

As Fontes de Pesquisas desta revisão foram escolhidas pela relevância e quantidade de trabalhos indexados contendo o tema proteção de dados, decisões automatizadas. As bibliotecas digitais escolhidas são: *IEEE Xplorer*[1], *Scopus*[2], *Web of Science*[3], *Google Acadêmico*[4], Periódicos Capes[5] e bases mantidas por universidades. Como exceção prevista no protocolo, inserções manuais também puderam ser utilizadas. Isto inclui trabalhos acadêmicos localizados em repositórios de universidades e revistas nacionais específicas sobre direito.

As *Strings* que montaram as expressões de busca foram retiradas da QP1, questão principal desta RSL. Após identificadas, permitiram cobrir com maior abrangência as buscas nas bases de dados escolhidas como fonte de coleta de trabalhos. Foram utilizados, como estratégia

[1]https://ieeexplore.ieee.org/Xplore/home.jsp
[2]https://www.scopus.com/home.uri
[3]https://login.webofknowledge.com
[4]https://scholar.google.com.br
[5]https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br

de busca, os motores de buscas das próprias bibliotecas digitais em seu modo avançado. As palavras utilizadas foram:

Tabela 4 – Termos de pesquisas e Sinônimos.

ID	Termos Principais	Termos Sinônimos/Traduzidos
1	*"Automatic Decisions"*	*"Automated Decisions", "Decisões Automatizadas"*
2	*"Data Protection"*	*"Personal Data Protection", "Information Protection", "Decisões Automatizadas", "Proteção de Dados", "Proteção de Dados Pessoais"*
3	*"Artificial Intelligence"*	*"Inteligência Artificial", IA*
4	*"Profiling"*	*"Perfilamento"*
5	*"General Data Protection Regulation"*	*"Regulamento Geral de Proteção de Dados", GDPR*
6	*"Lei Geral de Proteção de Dados"*	*LGPD*
7	*"Black Box Algorithm"*	

Fonte: Autor.

As palavras definidas na coluna "Termos Principais" detêm preferência em relação as palavras definidas na coluna "Termos Sinônimos/Traduzidos". As palavras de ID 1, 3 e 4 referem-se ao contexto específico de tratamento de dados que compõe o corte metodológico desta pesquisa. As palavras de ID 2, 5, 6 referem-se ao contexto geral e aponta para a área e instrumentos normativos essenciais para a pesquisa. O complemento ocorre com a palavra de ID 7, relacionada diretamente como o problema fático da pesquisa.

Com as *Strings* de buscas definidas, os sistemas internos de busca avançada de cada uma das bases de dados escolhida para a RSL foram alimentadas com elas em formato de expressões lógicas, como a tabela abaixo mostra.

Tabela 5 – Termos de pesquisas e Sinônimos.

ID	Expressão	Bases de Pesquisa
1	*TS=("personal data protection") OR TS= ("Information Protection") AND TS= ("automated decisions") AND TS= ("artificial intelligence") AND TS= ("non-discrimination") AND TS= ("GDPR") OR TS= ("General Data Protection Regulation")*	*Web of Science*

2	*((”personal data protection”) AND ”General Data Protection Regulation”) (((”General Data Protection Regulation”) OR ”GDPR”) AND ”artificial intelligence”) AND ”Black Box Algorithm”*	*IEEE*
3	*(ALL(”Automatic Decisions”) AND ALL (”Automatic Decisions”) AND ALL(gdpr))*	*Scopus*
4	*Qualquer campo contém Decisões Automatizadas E Qualquer campo contém ”Inteligência Artificial”E Qualquer campo contém ”LGPD”*	*Periódico Capes*
5	*”decisões autônomas”OR ”decisões automatizadas”AND ”proteção de dados”OR ”lei geral de proteção de dados”OR ”regulamento geral de proteção de dados”*	*Google Acadêmico*

Fonte: Autor.

Cada base de pesquisa teve a montagem da sua respectiva expressão. Embora a estrutura dessas expressões tenha sido alterada para atender o sistema de busca, o objetivo é o mesmo, trazer o máximo de trabalhos relacionados aos termos, de forma mais ampla possível.

A.2.3 Critérios de inclusão e exclusão

Os critérios de inclusão e exclusão desta RSL permitiram a identificação e a seleção dos trabalhos relevantes para adentrar a fase de extração de dados, reduzindo, também, a probabilidade de viés. Ressalta-se que o objetivo foi trazer para a revisão o máximo de trabalhos científicos relacionados com os termos de pesquisas.

Tabela 6 – Critérios de inclusão e exclusão.

ID	Critérios de inclusão	Critério de exclusão
1	Trabalhos que referenciaram diretamente no título o objeto da revisão.	Trabalhos anteriores a 2016.
2	Trabalho que referenciem diretamente no resumo o objeto da revisão.	Trabalhos que apenas citem e não discutam o objeto da revisão.
3	Trabalhos em inglês, Português e Espanhol.	Trabalhos que discutam somente aspectos tecnológicos do objeto da revisão.
4	Trabalhos entre 2016 e 2022.	Trabalhos não científicos.

5		Trabalhos em linguagem distinta da selecionada nesta revisão.
6		Trabalhos que não abordem o tema.

Fonte: Autor.

Em relação aos critérios de inclusão: 1 e 2 foram definidos para garantir que os trabalhos tenham ligação estreita com o objeto de estudo, enquanto o critério 3 foi definido com o propósito de expandir a revisão para além de trabalhos publicados no Brasil ou países de língua espanhola e o critério 4 foi definido para abarcar trabalhos que abordem os instrumentos normativos GDPR e LGPD.

Para os critérios de exclusão: 1 foi definido para garantir que os trabalhos abordem os principais instrumento de proteção de dados atuais que influenciam o Brasil (GDPR e LGPD). Os critérios 2, 3 e 6 foram definidos para garantir que exista uma discussão jurídica mais aprofundada nos trabalhos incluídos. Os critérios 4 e 5 tem por objetivo excluir trabalhos que não detenham uma rigidez científica relevante e que limitem, de certo modo, os trabalhos que tenham relevância quanto as linguagens (português para publicações locais, inglês e espanhol para publicação internacionais gerais).

A.2.4 Extração de dados

Os trabalhos selecionados para a fase de extração de dados foram submetidos aos seguintes critérios de extração:

- Informações dos Trabalhos: título do trabalho, lista de autores, tipos de estudo, ano de publicação e resumo.
- Os tipos de automatização de decisões são abordados.
- Os problemas enfrentados no estudo.
- As soluções propostas para enfrentamento de problemas com decisões automatizadas.

Os critérios definidos para extração de dados possibilitam a construção de informações relevantes acerca dos estudos. Estas informações permitem investigar a abrangência com o qual tema está sendo pesquisado, quais são os problemas enfrentados na literatura, quais são as propostas apresentadas nos estudos e se são satisfatórias. As datas de publicação permitem verificar a evolução cronológica dos trabalhos em termos de propostas e as informações dos trabalhos permitam uma visão geral dos estudos.

A.3 Avaliação de qualidade

A avaliação de qualidade (AQ) dos trabalhos selecionados nesta RSL foi realizada por intermédio de um sistema de pontuação com o propósito de avaliar a credibilidade, integridade

e relevância. Neste sentido, a avaliação de qualidade fora dividida em duas categorias, são elas: I) critérios gerais, adaptados de estudos existentes na literatura, cujo objetivo é avaliar tecnicamente a qualidade dos trabalhos selecionados - se apresentam objetivos claros, discussões gerais, se expõe explicitamente ameaças a validade e outras questões; II) critérios específicos, critérios designados de acordo com o escopo e questões de pesquisas desta RSL. As questões AQ1 até AQ7 foram retiradas da literatura, enquanto as demais foram baseadas no contexto desta revisão.

Tabela 7 – Questões para avaliação de qualidade.

#	**Questões**	**Possíveis respostas**
AQ1	Existe uma razão do porquê o estudo foi realizado? MAHDAVI-HEZAVEHI; GALSTER; AVGERIOU (2013)	Y = 1, N = 0, P = 0.5
AQ2	O documento é baseado em pesquisa (ou é meramente um relatório de "lições aprendidas" baseado na opinião de especialistas)? DYBÅ; DINGSØYR (2008)	Y = 1, N = 0, P = 0.5
AQ3	Existe uma declaração clara dos objetivos da pesquisa? DYBÅ; DINGSØYR (2008)	Y = 1, N = 0, P = 0.5
AQ4	Existe uma descrição adequada do contexto (indústria, ambiente de laboratório, produtos utilizados e assim por diante) em que a pesquisa foi realizada? (DYBÅ; DINGSØYR (2008) e MAHDAVI-HEZAVEHI; GALSTER; AVGERIOU (2013))	Y = 1, N = 0, P = 0.5
AQ5	O estudo é apoiado por uma ferramenta? DERMEVAL et al. (2014)	Y = 1, N = 0, P = 0.5
AQ6	As limitações deste estudo são explicitamente discutidas? DING et al. (2014)	Y = 1, N = 0, P = 0.5
AQ7	Existe uma discussão sobre os resultados do estudo? DERMEVAL et al. (2014)	Y = 1, N = 0, P = 0.5
AQ8	Apresenta a abordagem em que o estudo se apoia (estritamente jurídica, jurídica-tecnológica e tecnológica-jurídica)?	Y = 1, N = 0, P = 0.5
AQ9	Apresenta problema relevante referente ao Direito à explicação de Decisões automatizadas por IA?	Y = 1, N = 0, P = 0.5
AQ10	Apresenta propostas de solução para os problemas levantados?	Y = 1, N = 0, P = 0.5

Fonte: Autor.

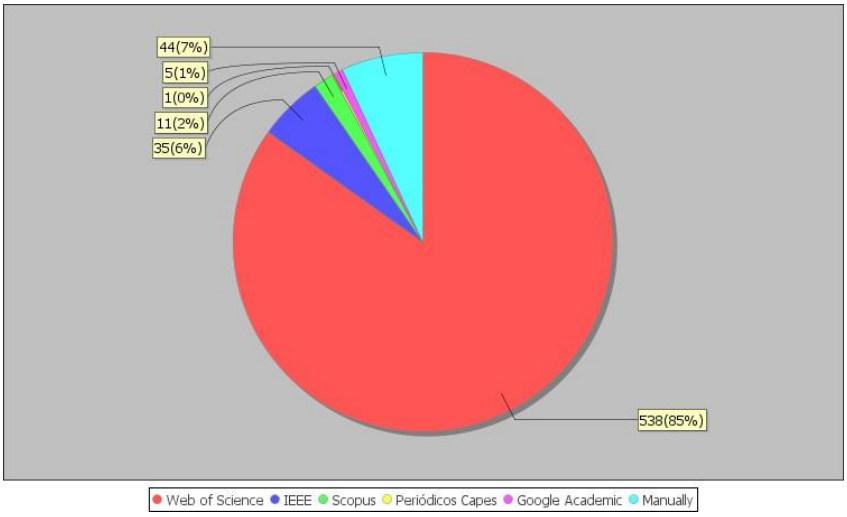
Todos os trabalhos aceitos foram submetidos a avaliação de qualidade, restando a nota

de corte acima de 70%, pois abaixo desta pontuação os trabalhos são considerados impróprios para discussão por deficiência em expor os critérios mínimos delimitados no Quadro 4.

A.4 Coleta e análise de dados

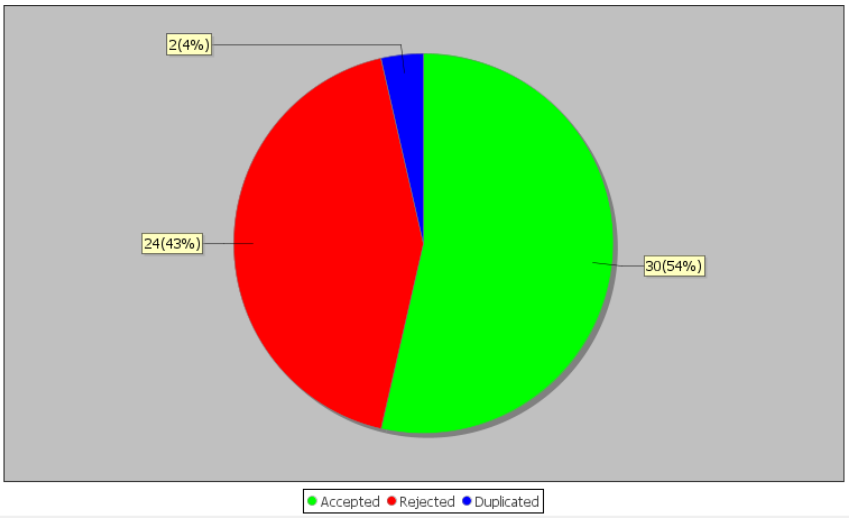
Como primeira etapa da execução da RSL, as *Strings* de Buscas foram utilizadas nos sistemas de buscas das bibliotecas digitais bases escolhidas, retornando como resultado, 556 referências. No intuito de ampliar o estudo, a técnica de *Snowballing* (análise das referências utilizadas em trabalhos científicos) foi aplicada, adicionando 44 referências para a revisão, totalizando 600 referências científicas. Em seguida, foram removidos os trabalhos duplicados através de análise realizada através da ferramenta StArt, resultando em remoção de 12 referências (2%). Em seguida, foram lidos 588 (98%) títulos, resumos e descrição das referências restantes, com a aplicação dos critérios de inclusão e exclusão definidos no protocolo, restando o total de 56 (9,33%) referências para a fase seguinte de extração de dados. A fase de extração de dados requer uma leitura mais aprofundada da referência. Das referências presentes na fase de extração de dados, 30 (53.57%) referências foram aprovadas para inclusão final na revisão, 24 (42.85%) das referências foram rejeitadas, 2 (3.57%) das referências foram avaliadas como duplicadas. Houve atualização na busca de trabalhos referentes aos anos de 2020, 2021 e 2022 totalizando em 34 trabalhos coletados, porém não foram classificados, portanto não fazem parte deste relatório. No todo, contanto com as atualizações, foram coletados 634 trabalhos. As Figuras 22 mostra a distribuição de trabalhos por fontes de coleta e a Figura 23 mostra a quantidade dos trabalhos aceitos até então para a extração de dados (inclusão final no trabalho).

Figura 22 – Fontes de trabalhos.



Fonte: o autor.

Figura 23 – Trabalhos selecionados para extração de dados.



Fonte: o autor.

A.5 Resultados

A Quadro 8 mostra todos os trabalhos incluídos na revisão sistemática e que foram submetidos a extração de dados. Os trabalhos podem ser visualizados graficamente pela Figura 25. A partir deste compilado de estudos, seguem os resultados de cada uma das perguntas levantadas nesta revisão.

Os trabalhos incluídos são compostos por buscas automatizadas nas fontes selecionadas no protocolo de pesquisas e de buscas diretas através das referências utilizadas nos trabalhos coletados nas fontes. Por esta razão, a partir da ID 20 da Tabela 8, se repetem alguns anos de publicação.

Tabela 8 – Trabalhos incluídos na RSL

ID	**Autor(es)**	**Ano de Publicação**
RSL1	KAMARINOU; MILLARD; SINGH	2016
RSL2	GOODMAN	2016
RSL3	DOSHI-VELEZ et al.	2017
RSL4	GOODMAN; FLAXMAN	2017
RSL5	SELBST; POWLES	2017
RSL6	MALGIERI; COMANDÉ	2017
RSL7	KINGSTON	2017
RSL8	EDWARDS; VEALE	2018
RSL9	KIM; ROUTLEDGE	2018
RSL10	BUTTERWORTH	2018
RSL11	CROCKETT; GOLTZ; GARRATT	2018
RSL12	BLACKLAWS	2018

RSL13	PESAPANE et al.	2018
RSL14	MONTEIRO	2018
RSL15	SINGH; COBBE; NORVAL	2019
RSL16	ZANATTA	2019
RSL17	LOPES	2019
RSL18	FELZMANN et al.	2019
RSL19	MOUGIAKOU; PAPADIMITRIOU; VIRVOU	2019
RSL20	LYNSKEY	2019
RSL21	KIM	2017
RSL22	EDWARDS; VEALE	2017
RSL23	WACHTER; MITTELSTADT; FLORIDI	2017
RSL24	WACHTER; MITTELSTADT; RUSSELL	2017
RSL25	KAMINSKI	2018
RSL26	SELBST; BAROCAS	2018
RSL27	MALGIERI	2019
RSL28	BIONI; LUCIANO	2019
RSL29	GUARDA	2019
RSL30	POLITOU; ALEPIS; PATSAKIS	2019

Fonte: o autor.

A Figura 24 mostra visualmente os trabalhos agrupados por ano de publicação.

Figura 24 – Gráfico de publicação por ano.

Fonte: o autor.

A.5.1 Resultados Questão de Pesquisa 1

QP1: Como os pesquisadores estão abordando o direito à explicação de decisões automatizadas realizados por algoritmos de IA?

De início, toda a execução desta RSL fornece um indicativo de como os pesquisados estão abordando o direito à explicação em decisões automatizadas realizadas por algoritmos complexos (Inteligência Artificial). Dos critérios escolhidos para as buscas dos trabalhos, levando-se em consideração as bases digitais e inserções manuais de referências, somente 600 referências foram coletados. As abordagem trazidas pelos 30 estudos incluídos nesta RSL foram agrupados em categorias: I) abordagem jurídica, onde predominantemente o aspecto jurídico-normativo foi levado em consideração, não havendo discussão tecnológica; II) abordagem jurídico-tecnológica, onde ambos os aspectos foram levados em consideração para o desenvolvimento do trabalho e; III) abordagem tecnológica-jurídica, onde predominantemente o aspecto tecnológico foi considerado, restando a discussão jurídica como suporte argumentativo.

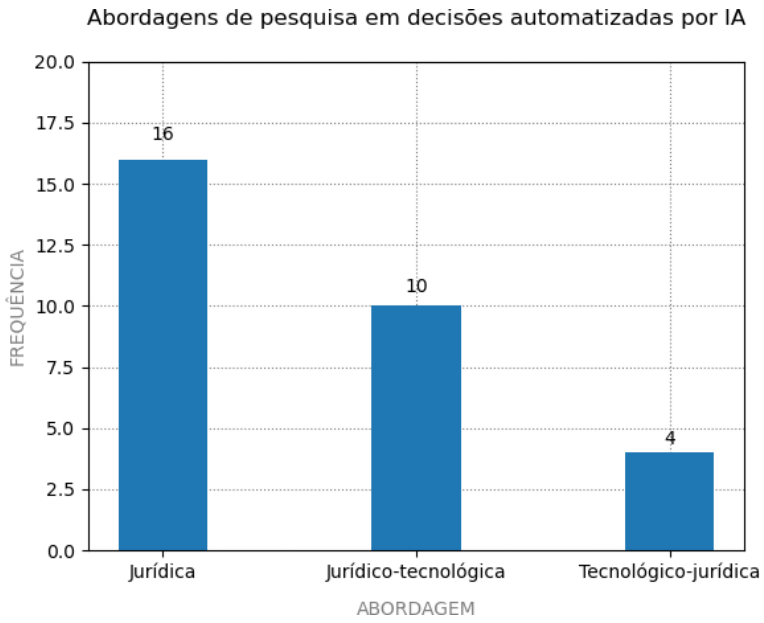
Tabela 9 – Abordagem dos trabalhos incluídos.

Abordagem	**Trabalhos**	**Frequência**	**%**
Abordagem jurídica	RSL1, RSL2, RSL4, RSL11, RSL12, RSL14,RSL15, RSL16, RSL17, RSL20, RSL22, RSL23, RSL24, RSL25, RSL27, RSL28	16	53.33 %
Abordagem jurídico-tecnológica	RSL3, RSL5, RSL6, RSL7, RSL9, RSL10, RSL13, RSL18, RSL19, RSL30	10	33.33 %
Abordagem tecnológica-jurídica	RSL8, RSL21, RSL26, RSL29	4	13.33 %

Fonte: o autor.

A Figura 25 mostra visualmente a disposição das frequências de trabalhos por tipos de abordagens.

Figura 25 – Gráfico de abordagens de pesquisa.



Fonte: o autor.

Os resultados, sistematizados na Tabela 9, assim como podem ser visualizados graficamente na Figura 25, indicam que uma abordagem somente jurídica é predominante (53.33%). Igualmente, uma abordagem mista com predominância jurídica mostra-se significativa (33.33%). Isto sugere que existe uma forte movimentação para um estudo interdisciplinar envolvendo tanto a área jurídica quanto a área da tecnologia da informação.

A.5.2 Resultados Questão de Pesquisa 2

QP2: Quais problemas enfrentados relacionados ao direito à explicação em decisões automatizadas?

Os problemas levantados pelos estudos incluídos nesta RL, para melhor entendimento, foram agrupados em três categorias - as categorias foram definidas com base nos principais problemas enfrentados nos trabalhos - a saber: I) opacidade algorítmica, voltados aos problemas de complexidade, tornando ininteligível o processo de tratamento de dados; II) discriminação algorítmica, voltados aos problemas de viés discriminatório no tratamento de dados, incluindo-se em processos judiciais; III) adequação e *compliance*, voltados aos problemas de conformidade com normas jurídicas e regulatórias.

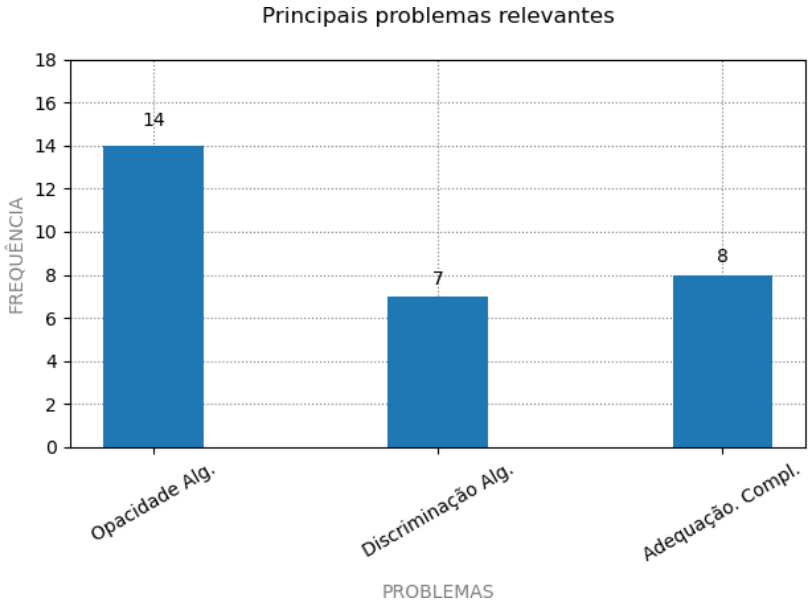
Tabela 10 – Problemas trazidos pela literatura.

Problemas	**Trabalhos**	**Frequência**	**%**
Opacidade algorítmica	RSL1, RSL3, RSL5, RSL6, RSL8, RSL9, RSL11, RSL14, RSL15, RSL16, RSL17, RSL18, RSL22, RSL24, RSL25	14	46.66 %
Discriminação algorítmica	RSL2, RSL4, RSL10, RSL12, RSL20, RSL21, RSL28	7	23.33 %
Adequação e *compliance*	RSL7, RSL13, RSL19, RSL23, RSL26, RSL27, RSL29, RSL30	8	26.66 %

Fonte: o autor.

A Figura 25 mostra visualmente a disposição das frequências de trabalhos por problemas de pesquisa.

Figura 26 – Gráfico de problemas relevantes de pesquisa.



Fonte: o autor.

Os resultados, sistematizados na Tabela 10, assim como podem ser visualizados graficamente na Figura 26, sugerem uma considerável concentração de trabalhos que enfrentam o problema de opacidade algorítmica (46.66%). No entanto, o problema da discriminação algorítmica é enfrentado em 7 trabalhos (23.33%). Isto pode indicar que a opacidade algorítmica detém importância para o desenvolvimento de soluções para o problema de pesquisa abordado neste pré-projeto.

A.5.3 Resultados Questão de Pesquisa 3

QP3: Quais as soluções propostas para o enfrentamento dos problemas levantados?

As propostas de solução trazidas pelos estudos incluídos na revisão foram agrupadas em cinco categorias - as categorias foram definidas com base nas principais propostas encontradas nos trabalhos. São elas: I) construção normativa técnica, cujo foco é a implementação de algoritmos baseados em privacidade de dados; II) aplicação do princípio da transparência, onde o conceito de transparência de processamento é considerado suficiente para implementação de ferramentas explicativas; III) metodologias múltiplas, os trabalhos propõem um conjunto de ferramentas normativas e técnicas para efetivação do direito à explicação; IV) Interpretação normativa, é proposto a expansão interpretativa de forma flexível e funcional dos direitos à explicação focada em informações significativas e; V) Não Especificado - Diversos, diz repeito a propostas que não se enquadram nas categorias anteriores.

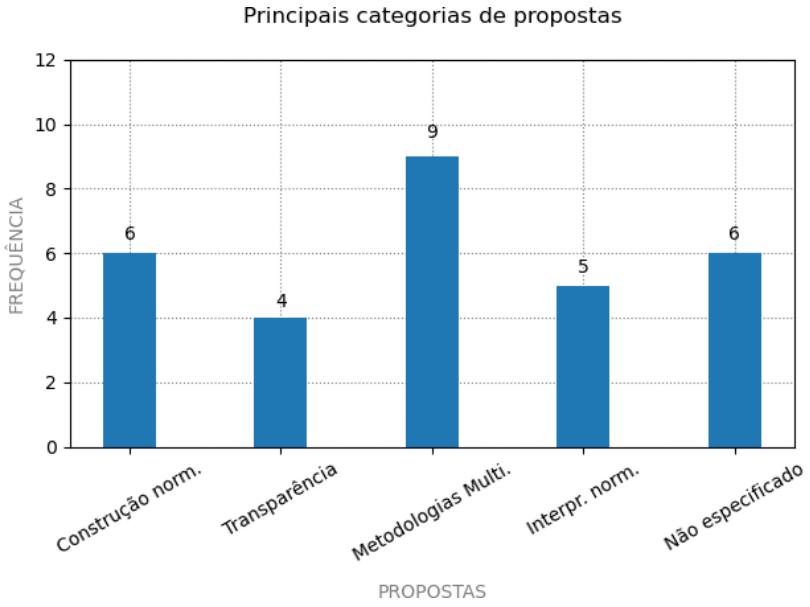
Tabela 11 – Propostas de solução trazidos pela literatura.

Propostas	**Trabalhos**	**Frequência**	**%**
Construção normativa técnica	RSL4, RSL14, RSL19, RSL23, RSL24, RSL27	6	20 %
Aplicação do princípio da transparência	RSL1, RSL6, RSL16, RSL18	4	13.33 %
Metodologias múltiplas	RSL2, RSL8, RSL11, RSL13, RSL17, RSL21, RSL22, RSL26, RSL30	9	30 %
Interpretação normativa	RSL3, RSL5, RSL15, RSL25, RSL28	5	16.66 %
Não Especificado	RSL7, RSL9, RSL10, RSL12, RSL20, RSL29	6	20 %

Fonte: o autor.

A Figura 27 mostra visualmente a disposição das frequências de trabalhos por problemas de pesquisa.

Figura 27 – Gráfico de propostas relevantes de pesquisa.



Fonte: o autor.

Os resultado, sistematizados na Tabela 11, assim como podem ser visualizados graficamente na Figura 27, indicam que a aplicação do princípio da transparência, com 4 trabalhos (13.33%), e a aplicação de metodologias múltiplas, com 9 trabalhos (30%). Desconsiderando os trabalhos que não houve uma indicação específica, existe um certo equilíbrio entre as propostas de solução no enfrentamento do problema.

A.6 Análise de literatura e discussão

Esta análise de literatura tem como objetivo o confronto dos trabalhos incluídos da RSL, referentes ao objeto, problema e contexto deste projeto de pesquisa, com a literatura complementar, que faz ligação direta com a privacidade e proteção de dados pessoais.

Os trabalhos incluídos na revisão sistemática podem indicar que o objeto e contexto de estudo ora proposto - a privacidade, proteção de dados e decisões automatizadas - não contém maturidade suficiente devido ao número reduzido de publicações relevantes até a finalização desta revisão. Uma abordagem estritamente jurídica, apesar de ser predominante, não é a única abordagem em destaque trazida pela literatura. Abordagens jurídica-tecnológica e tecnológicas-jurídicas, ambas com aspectos interdisciplinares envolvendo a área do Direito e da tecnologia da informação e comunicação, possuem relevância para o estudo do tema, além de somarem 46.66% dos trabalhos incluídos na revisão. Isto pode indicar uma tendência, nos estudos, em considerar aspectos tecnológicos para propostas de solução destinadas à resolução de problemas de tomada de decisão automatizada.

Dentre os trabalhos incluídos, foi possível classificá-los quanto aos problemas que enfrentaram. Os problemas com opacidade algorítmica, apesar de serem caracterizados como

problemas tecnológicos, afetam diretamente o exercício do direito à explicação de decisões automatizadas em contexto de Inteligência Artificial. Um algoritmo pode ser descrito como uma sequência de instruções lógicas, sequencialmente organizadas capazes de resolver problemas determinados (EDWARDS; VEALE, 2017). Todo software tem seu planejamento estabelecido através da criação destes algoritmos. No entanto, por si só, estes mesmos algoritmos não podem funcionar como um programa, pois se faz necessário sua tradução para uma linguagem de programação, formando o código fonte do programa. Estes códigos, então, são convertidos em linguagem de máquina, constituindo o software capaz de ser executado e compreendido por dispositivos informáticos. Em EDWARDS; VEALE (2017) explica que a opacidade de um algoritmo tem o efeito oposto do determinado pela necessidade de transparência de operações automatizas, sobretudo realizadas por Inteligência Artificial através de aprendizado de máquina. Isto porque estes modelos algorítmicos possuem lógica de processamento imprevisíveis, o que pode resultar na impossibilidade de explicação do processo como um todo e em cenário de decisão específica.

A discriminação algorítmica, por sua vez, pode ser resultado de inferência sobre dados enviesados (GOODMAN, 2016). A construção das bases de dados são fundamentais para utilização de sistemas automatizados, em especial por sistemas que utilizam técnicas de Inteligência Artificial. Esta revisão aponta que 23.33% dos trabalhos abordam o problema de discriminação algorítmica, indicando preocupação, por parte da literatura, com os dados e não somente com o software que os utilizam. Problemas com adequação e *compliance* também são levados em consideração pela literatura. Nesta revisão 26.66% dos trabalhos incluídos são direcionados ao enfrentamento deste problemas. Em conjunto, ambos refletem 49.99% dos problemas trazidos pela literatura. A adequação envolve a observâncias das legislações e regulamentos sobre proteção de dados e decisões automatizadas.

Quanto às propostas dos trabalhos, não houve uma convergência para uma solução genérica ou sequer uma predominância entre eles. Das categorias estabelecidas para sistematização dos resultados, soluções baseadas em metodologias múltiplas - aquelas que agregam tanto ferramentas normativas quanto técnicas - detém o destaque mais significativo entre todas as elencadas com 30%. Esta proposta corrobora com um posicionamento interdisciplinar de uma abordagem de pesquisa envolvendo o Direito e a tecnologia da informação, embora não seja possível estabelecer uma correlação entre este tipo de proposta e a uma abordagem mista de pesquisa. A solução baseada na aplicação do princípio da transparência se encontra no quarto lugar, com 13.33%. Este resultado pode significar que a transparência, de forma isolada, não é capaz de lidar com problemas de decisões automatizadas por Inteligência Artificial. Dentre soluções de interpretação normativa e construção normativa técnica, ambas com somatório de 36.66%, não representam maioria das propostas apresentadas. Destaca-se que em 20% dos trabalhos não mostram solução compatível com nenhuma categoria estabelecida para esta sistematização.

Dados e informações são constantemente utilizados para representação de indivíduos,

sendo irrelevante o suporte físico do qual se faz uso para sua persistência ou tratamento. O fluxo informacional evolui com o aperfeiçoamento das tecnologias de informação e comunicação, em especial com a computacional. Em inúmeros modelos de negócio, os dados têm papel fundamental para alavancar o sucesso desses empreendimentos e isto pode caracterizar um risco significativo a direitos fundamentais, como a privacidade, saúde, vida e outros (MENDONÇA, 2014). Estes dados, por sua vez, são repassados para terceiros sem ao menos o conhecimento do titular do dado, o que se configura como violação ao direito à privacidade.

A privacidade tem sido utilizada como ponto inicial para discussões sobre proteção de dados pessoais e sua regulamentação. No entanto, existe uma movimentação em prol do reconhecimento do direito à proteção de dados pessoais como um direito fundamental, pois reside no tema peculiaridades que trazem certa autonomia valorativa suficiente para garantir o status de direito constitucionalmente assegurando (SENADO FEDERAL, 2019). Isto é proposto pela PEC 17/2019:

> Art. 1º Inclua-se no art. 5º, da Constituição Federal, o seguinte inciso XII-A:
>
> "Art. 5º .....
>
> XII-A - é assegurando, nos termos da lei, o direito à proteção de dados pessoais, inclusive nos meios digitais.
>
> [...]

A PEC ainda traz em sua justificativa o mesmo movimento em relação a outros países, como Chile, Portugal, Estônia e Polônia. Além destes países, legislações específicas sobre proteção de dados se espalham pelo restante do planeta, impulsionadas pelo uso desenfreado de dados pessoais em ambiente digital. Trata-se, portanto, de se evidenciar que a proteção dados pessoais detém significado relevante, especialmente quando sistemas de IA decidir sobre o destino do titular de dados de forma indiscriminada.

A.6.1 Síntese cronológica da literatura

Inicia-se neste ponto uma análise cronológica sobre os principais trabalhos incluídos neste revisão bibliográfica em principais pontos relevantes.

2016: No trabalho de GOODMAN são considerados duas principais ferramentas principiológicas para lidar com discriminação algorítmica - sanitização dos dados e transparência algorítmica. A legislação também abre caminho para inspeções de algoritmos por terceiros ou "auditorias de algoritmos". As auditorias de algoritmos suportadas pelo Regulamento Geral de Proteção de Dados podem desempenhar um papel crítico para tornar os algoritmos menos discriminatórios e mais responsáveis. O KAMARINOU; MILLARD; SINGH, fornece uma breve visão geral de algumas das considerações e desafios relacionados ao uso de modelos de aprendizado de máquina para processar dados pessoais de indivíduos sob o Regulamento Geral de Proteção de Dados da União Europeia.

2017: Em KIM é abordado como a legislação especializada pode se utilizar de auditorias algorítmicas para detectar e corrigir problemas de viés discriminatório. No trabalho de WACHTER; MITTELSTADT; FLORIDI é proposto uma série de medidas legislativas e políticas que, se tomadas, podem melhorar a transparência e a responsabilidade da tomada de decisão automatizada no contexto da Regulamento Geral de Proteção de Dados. Em outro relevante estudo de WACHTER; MITTELSTADT; RUSSELL são propostos objetivos para explicação destinados ao titular de dados: informar e ajudar na compreensão de determinada decisão; fornecer fundamentos para contestação e: compreender o que pode ser alterado para mudança de resultados futuros, superando limitações trazidas por legislação e regulamentos de proteção de dados em contexto de decisões automatizadas. No estudo de MALGIERI; COMANDÉ é argumentado que uma interpretação sistêmica é necessária para trazer benefícios para os indivíduos e para as instituições. Em DOSHI-VELEZ et al., a explicação é focada como ferramenta para prevenção de consequências negativas dos sistemas de Inteligência Artificial. No trabalho de GOODMAN; FLAXMAN é sumarizado o potencial impacto da Regulamento Geral de Proteção de Dados no uso rotineiro de algoritmos de aprendizagem de máquina. No estudo de EDWARDS; VEALE é discutido o direito a explicação no contexto da Regulamento Geral de Proteção de Dados e sua insuficiência para evitar danos algorítmicos, uma vez que existe uma restrição na legislação sobre o tema. No trabalho de SELBST; POWLES, é defendido o direito à explicação na forma de informações significativas para o titular dos dados tratados. O debate do "direito à explicação", versa sobre como é complicado, complexo e um desafio técnico não trivial em aproveitar todo o poder dos sistemas Inteligência Artificial enquanto opera com lógica interpretável para os seres humanos. Por último, no trabalho de KINGSTON considera o *compliance* como fundamental em relação à legislação de proteção de dados, em especial sobre decisões automatizadas e propõe quadro área de conformidade para adequação em contexto da Regulamento Geral de Proteção de Dados.

2018: Em KAMINSKI, há discussão sobre a limitação dos regulamentos sobre responsabilização algorítmica, incluindo-se o direito à explicação de decisões automatizadas sobre Inteligência Artificial. Em relação ao trabalho de BUTTERWORTH, uma discussão sobre justiça algorítmica é trazida no contexto de Inteligência Artificial para evitar consequências socialmente negativas. No estudo EDWARDS; VEALE existe preocupação com decisões automatizadas por Inteligência Artificial no contexto empresarial e de governo. No que tange o trabalho de KIM; ROUTLEDGE, é evidenciada a necessidade em concretizar o direito à explicação de decisões automatizadas. O estudo de CROCKETT; GOLTZ; GARRATT discute o impacto do Regulamento Geral de Proteção de Dados em projetos de pesquisa que contêm elementos de inteligência computacional realizados dentro de uma universidade ou com um parceiro acadêmico.

O estudo BLACKLAWS, explora as questões de responsabilidade e transparência em relação ao uso crescente de algoritmos de aprendizado de máquina. Em PESAPANE et al. consideram-se questões de responsabilidade, tanto legais quanto éticas no contexto médico. Dispositivos médicos são amplamente utilizados e imprevisíveis. Por isto, o trabalho defende uma regulamentação da utilização de sistemas de Inteligência Artificial para aplicações médicas. No estudo de SELBST; BAROCAS há uma análise de técnicas de Inteligência Artificial para viabilização do direito à explicação e aponta falhas nos modelos atuais de Inteligência Artificial para esta explicação no processamento lógico. No trabalho de MONTEIRO, é realizada uma análise dos aspectos que são jurídicos de proteção de dados especificamente sobre direito à explicação em contexto de decisões automatizadas no Brasil.

2019: Em início, se tem o trabalho de MALGIERI. O objetivo deste artigo é analisar as leis nacionais dos Estados-Membros recentemente aprovadas que implementaram o Regulamento Geral de Proteção de Dados no campo da tomada de decisão automatizada (proibição, exceções, salvaguardas). Em SINGH; COBBE; NORVAL, argumenta-se que métodos de proveniência de dados são promissores como meio para implementar transparência e garantir uma explicação significativa para o titular do dado pessoal. Em LOPES, destina-se a enfrentar problemas de concretização do direito à explicação no contexto da Regulamento Geral de Proteção de Dados, trazendo a tomada de decisão para o âmbito de sistemas e referente à uma decisão específica. No trabalho de FELZMANN et al., aborda-se o tema de transparência em Inteligência Artificial, integrando aspectos legais, sociais e éticos. O estudo de ZANATTA, encontra-se discussão sobre perfilamento, discriminação algorítmica e faz uma relação com Código de Defesa do Consumidor e a Lei Geral de Proteção de Dados. Em GUARDA, discutem-se a tomada de decisão por Inteligência Artificial no contexto da saúde, em particular atenção às questões sobre proteção de dados pessoais. No trabalho de MOUGIAKOU; PAPADIMITRIOU; VIRVOU há uma combinação entre Informática com Ciência Jurídica, no mapeamento de áreas específicas de conflito entre aplicativos educacionais personalizados com técnicas de Inteligência Artificial e o Regulamento Geral de Proteção de Dados, abordando deficiências e fornecendo soluções aplicáveis, na forma de diretrizes para desenvolvedores. No estudo de LYNSKEY examina-se a aplicação das iterações entre a Regulamento Geral de Proteção de Dados e Diretiva de Aplicação da Lei e na jurisprudência do Tribunal de Justiça da União Europeia. No artigo de POLITOU; ALEPIS; PATSAKIS, é apresentada relação entre IA e o princípio da transparência em contexto de Big Data, no que tange a formação de perfis e tomada de decisão automatizadas. Por último neste ano, o artigo de BIONI; LUCIANO aborda o princípio da precaução para regulamentação de Inteligência Artificial, levando-se em consideração as leis de proteção de dados pessoas.

Observa-se que os esforços estão voltados ao combate de problemas relacionados à decisões automatizadas por IA em diversas vertentes e abordagens. Em todos os estudos, seguem análise normativa, em especial do artigo 20, §1º da lei 13709/2018 e no Regulamento de Proteção de Dados Pessoais Europeu, e análise tecnológica, investigando a possibilidade de se extrair uma explicação significativa de processos utilizados por sistemas inteligentes. Apesar destes esforços, não há um posicionamento capaz de trazer suficiência para a questão.

A.7 Ameaças à validade

Esta seção descreve preocupações que devem ser consideradas para reproduções futuras deste estudo e outros aspectos que devem ser pesados para generalizar os resultados desta RSL. Na organização desta seção, as ameaças à validade foram classificadas em categorias Interna e Externa.

Sobre a validade interna, algumas decisões de cunho subjetivo podem ter ocorrido durante as análises e extração de dados realizadas nesta RSL, uma vez que nem todos os estudos coletados apresentaram elementos suficientemente claros, como descrição, objetivos ou resultados adequados, o que dificultou uma análise objetiva quanto à aplicação dos critérios de inclusão de trabalhos ou extração imparcial de dados. Com o intuito de minimizar este aspecto e atenuar as ameaças advindas de um possível viés na compreensão dos estudos, o processo de seleção dos trabalhos foi realizado mediante dupla análise dos requisitos definidos do protocolo.

A validade externa preocupa-se com a generalização dos resultados da RSL e está relacionada diretamente com o grau de representatividade dos estudos para cada ponto revisado. A fim de reduzir essas ameaças externas, o processo de definição de *Strings* para a coleta de trabalhos foi definido após várias pesquisas de teste, sendo validadas aquelas que trouxeram, de forma mais ampla, o maior número de trabalhos relacionados. Outro ponto relevante está na escolha das bases de dados escolhidas.

Testes foram realizados para saber da existência de material disponível que se relacione com o tema, sendo incluídas aquelas que retornaram resultados positivos. É possível que alguns estudos tenham ficado fora do escopo final desta revisão, pois não houve acesso ao seu conteúdo. Assim, com o objetivo de minimizar esta ameaça, entrou-se em contato com os autores, através de e-mail, solicitando os documentos, porém até a conclusão desta revisão não se obteve resposta.

Sobre a validade do conteúdo, os trabalhos selecionados têm uma limitação temporal até o ano de 2019. A partir de então, os trabalhos até o ano de 2022 foram coletados, mas não classificados e analisados até o presente momento. Porém, como forma de mitigação a atualização será realizada até a defesa da dissertação de mestrado.

A.8 Observações finais

A revisão sistemática realizada com este protocolo aborda a privacidade, a proteção de dados e as decisões automatizadas com sistemas de Inteligência Artificial. A revisão indica uma falta de maturidade sobre os temas, com uma predominância de abordagens jurídicas, mas também com abordagens interdisciplinares que consideram a tecnologia. Os principais problemas identificados incluem opacidade algorítmica, discriminação algorítmica, adequação e compliance. As propostas para solucionar esses problemas também variam bastante, com soluções baseadas em metodologias múltiplas sendo as mais comuns. Nesta revisão é possível apontar a importância dos dados em modelos de negócio e concluindo que há uma necessidade crescente de lidar com questões éticas e legais em torno da Inteligência Artificial e dos dados.

APÊNDICE B – PROTOCOLO REVISÃO - DADOS E INTELIGÊNCIA ARTIFICIAL

A revisão de literatura narrativa (RLN) se dará a partir deste protocolo de revisão simplificado. Este protocolo tem como objetivo direcionar a busca por estudos e trabalhos referentes à IA e sua relação com dados. Para este fim, os métodos de buscas, as palavras-chaves, os tipos de trabalhos admitidos, os critérios para inclusão e os critérios para extração de dados serão indicados nesta RLN.

Além dos objetivos descritos no parágrafo anterior, cabe indicar a delimitação que será adotada. Isto porque a área de Inteligência Artificial tem uma quantidade significativa de pesquisas e trabalhos publicados disponíveis. Desta forma, a revisão será guiada por dois aspectos de pesquisas: a técnica da tecnologia digital e a do Direito.

B.1 Fontes de pesquisa, tipos de trabalhos e Strings de busca

As Fontes de Pesquisas desta revisão foram escolhidas pela relevância e quantidade de trabalhos indexados contendo o tema de Inteligência Artificial e sua relação com os dados. As bibliotecas digitais escolhidas são: *IEEE Xplorer*[1], *ACM Digital Libery*[2], *Scopus*[3], *Web of Science*[4], *Google Acadêmico*[5], Periódicos Capes[6] e bases mantidas por universidades. As bases *IEEE Xplorer*, *ACM Digital Libery*, *Web of Science* são as principais fontes para busca dos trabalhos com aspectos tecnológicos. As bases *Scopus*, *Google Acadêmico*, Periódicos Capes e as bases mantidas por universidades foram utilizadas como forma complementar das buscas e com enfoque na área do Direito. Desta forma, os tipos de trabalhos definidos para adentrar na RLN são:

Tabela 12 – Tipos de trabalhos aceitos.

ID	**Tipos de trabalhos**	
1	Artigos científicos	*Revistas científicas indexadas*
2	Artigos científicos	*Conferência indexados*
3	Teses	*Literatura cinzenta*
4	Dissertações	*Literatura cinzenta*
5	Monografias	*Literatura cinzenta*
6	Livros	*Literatura cinzenta*

Fonte: Autor.

[1]https://ieeexplore.ieee.org/Xplore/home.jsp
[2]https://dl.acm.org/
[3]https://www.scopus.com/home.uri
[4]https://login.webofknowledge.com
[5]https://scholar.google.com.br
[6]https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br

As buscas foram realizadas através dos sistemas de buscas automatizados fornecidos por cada uma das bibliotecas digitais. Assim sendo, se fez necessário estabelecer as palavras chaves utilizadas nesses motores. Buscas manuais foram utilizadas como forma de ampliação das buscas e isto foi realizado tanto para o aspecto jurídico quanto para o aspecto tecnológico. Nos caso das buscas manuais, a metodologia será realizada por buscas nos sumários dos trabalhos e em caso de inexistência de sumário, a busca fora realizada pelas seções. As palavras-chave (*"Strings"*) são:

Tabela 13 – Termos de pesquisas e Sinônimos.

ID	Termos Principais	Termos Sinônimos/Traduzidos/Secundários
1	*"Artificial Intelligence"*	*"Inteligência Artificial", IA*
2	*"Machine Learning"*	*"Deep Learning"*, "Aprendizado de Máquina", "Aprendizado Profundo"
3	*"Big Data"*	"Dado", "Informação", "Conhecimento"
4	*"Black Box Algorithm"*	"Opacidade algorítmica", "Algoritmo opaco"
5	*"Direitos Fundamentais"*	*"Garantias Fundamentais"*
6	*"Responsabilidade Civil"*	
7	*"Transparência", "Níveis de Transparência"*	"graus de transparência"
8	*"Explicação", "Explicabilidade"*	"Características da explicação", "Tipos de explicação"

Fonte: Autor.

Os termos principais 1, 2 e 3 têm por objetivo localizar os trabalhos de forma ampla para fornecer conhecimentos basilares sobre os temas fundamentais do projeto de pesquisa. Com as Strings de buscas definidas, os sistemas internos de busca avançada de cada uma das bases de dados escolhida para a RLN foram alimentadas com elas em formato de expressões lógicas, como a tabela abaixo mostra.

Tabela 14 – Termos de pesquisas e Sinônimos.

ID	Expressão	Bases de Pesquisa
1	Combinação dos termos com buscas automatizadas simples	*IEEE Xplorer*
2	Combinação dos termos com buscas automatizadas simples	*ACM Digital Libery*
3	Combinação dos termos com buscas automatizadas simples	*Scopus*
4	Combinação dos termos com buscas automatizadas simples	*Web of Science*

5	Combinação dos termos com buscas automatizadas simples	*Google Acadêmico*
6	Combinação dos termos com buscas automatizadas simples	*Periódico Capes*

Fonte: Autor.

Cada base de pesquisa teve a montagem da sua respectiva expressão. Embora a estrutura dessas expressões tenha sido alterada para atender o sistema de busca, o objetivo é o mesmo, recuperar os trabalhos relevantes para o projeto de pesquisa.

B.2 Critérios de inclusão

Os critérios de inclusão desta RLN permitiram a identificação e a seleção dos trabalhos relevantes para adentrar a fase de extração de dados. Seguem os critérios de inclusão adotados. Cabe ressaltar, que os critérios apresentados não não exaustivos, servem apenas como guias.

Tabela 15 – Critérios de inclusão e exclusão.

ID	**Critérios de inclusão**
1	Trabalhos que referenciaram diretamente no título o objeto da revisão.
2	Trabalho que referenciem diretamente no resumo o objeto da revisão.
3	Trabalhos em inglês, Português e Espanhol.
4	Trabalhos entre 2013 e 2023.

Fonte: Autor.

Os trabalhos adicionados na Revisão Sistemática de Literatura que fundamenta a principal discussão deste projeto de pesquisa podem ser utilizados nesta revisão narrativa de literatura, caso exista relação próxima com o objetivo traçado.

B.3 Extração de dados

Os trabalhos selecionados para a fase de extração de dados foram submetidos aos seguintes critérios de extração:

- Conceitos de ambiente digital e algoritmos.
- Os conceitos e definições sobre Inteligência Artificial.
- Tipos, paradigmas e aspectos de Inteligência Artificial.
- Aprendizado de máquina e aprendizagem profunda.
- Autonomia tecnológica.
- Direitos e garantias fundamentais e sua relação com Inteligência Artificial.

- Relação entre direito à explicação e o princípio da motivação de decisões judiciais.
- Conceitos de transparência, níveis de transparência e características da transparência.
- Conceitos de explicabilidade, tipos de explicação e características da explicação.
- Quaisquer outros conceitos, dados e características relevantes para a pesquisa.

Os critérios definidos para extração de dados possibilitam a construção de informações relevantes acerca dos estudos.