UNIVERSIDADE FEDERAL FLUMINENSE
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO CONSTITUCIONAL
FACULDADE DE DIREITO

THAÍS MOURA CARREIRA

INTELIGÊNCIA ARTIFICIAL E SUA APLICABILIDADE AO DIREITO BRASILEIRO: UMA ABORDAGEM AO MUNDO JURÍDICO

Niterói
2023

THAÍS MOURA CARREIRA

INTELIGÊNCIA ARTIFICIAL E SUA APLICABILIDADE AO DIREITO BRASILEIRO: UMA ABORDAGEM AO MUNDO JURÍDICO

Dissertação apresentada ao Programa de Pós-Graduação em Direito Constitucional da Universidade Federal Fluminense, como requisito parcial à obtenção do título de Mestre em Direito Constitucional.

Linha de Pesquisa: Teoria e História do Direito Constitucional e Direito Constitucional Internacional e Comparado.

Orientadora: Prof.ª Dr.ª Clarissa Maria Beatriz Brandão de Carvalho Kowarski

Niterói

2023

Ficha catalográfica automática - SDC/BFD
Gerada com informações fornecidas pelo autor

C314i Carreira, Thais Moura
INTELIGÊNCIA ARTIFICIAL E SUA APLICABILIDADE AO DIREITO BRASILEIRO: UMA ABORDAGEM AO MUNDO JURÍDICO / Thais Moura Carreira. - 2023.
119 f.

Orientador: Clarissa Brandão.
Dissertação (mestrado)-Universidade Federal Fluminense, Faculdade de Direito, Niterói, 2023.

1. Inteligência artificial. 2. Sistema. 3. Responsabilidade. 4. Inovação. 5. Produção intelectual. I. Brandão, Clarissa, orientadora. II. Universidade Federal Fluminense. Faculdade de Direito. III. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

THAÍS MOURA CARREIRA

INTELIGÊNCIA ARTIFICIAL E SUA APLICABILIDADE AO DIREITO

Dissertação apresentada ao Programa de Pós-Graduação em Direito Constitucional da Universidade Federal Fluminense, como requisito parcial à obtenção do título de Mestre em Direito Constitucional.

Aprovada em 11 de Dezembro de 2023

BANCA EXAMINADORA

_____

Prof.ª Drª. Clarissa Maria Beatriz Brandão de Carvalho Kowarski (Orientadora)
Universidade Federal Fluminense (UFF)

_____

Prof. Dr. Leonardo da Silva Sant'anna
Universidade do Estado do Rio de Janeiro (UERJ)

_____

Prof. Dr. Marco Aurelio Casamasso
Universidade Federal Fluminense (UFF)

Aos amigos, familiares e namorado pelo incentivo e apoio constantes. Á minha orientadora por estar aberta a trocar conhecimentos e por buscar sempre a expansão dos horizontes de seus alunos.

AGRADECIMENTOS

Desde o início da minha vida, meus pais, Joselita e Alexandre, sempre foram meus maiores incentivadores. Foi graças a eles que tive acesso a uma boa educação, que somado a um ambiente acolhedor, saudável e repleto de amor, fez com que atualmente eu tivesse condição de concluir o Mestrado em Direito Constitucional pela Universidade Federal Fluminense.

As lições de resiliência e os constantes incentivos recebidos dos meus pais, com certeza, foram os pontos chaves para que fosse possível chegar até esse momento. Sendo assim, agradeço a eles por todo apoio, pelo amor e por toda a dedicação a mim e à minha educação.

Além deles, meu namorado Nicholas foi de extrema importância, já que suportou todos os finais de semana dedicados ao mestrado e à pesquisa, sendo meu porto seguro nos momentos difíceis e nunca me deixando duvidar de que seria possível concluir o curso. Com ele aprendi muito sobre dedicação e amor, ao mesmo tempo que aprendia sobre inteligência artificial e os impactos positivos na sociedade.

Minha avó Izabel e minhas tias-avós Maria, Alice e Bernadete também merecem destaque. Ter a presença e o carinho delas durante a infância, adolescência e a vida adulta e ver o orgulho que elas tinham do meu crescimento me deu ainda mais forças para seguir em frente e conquistando ainda mais vitórias.

“Para ser grande, sê inteiro: nada
Teu exagera ou exclui”

(Fernando Pessoa)

RESUMO

No cenário jurídico contemporâneo, a crescente interseção entre tecnologia e prática jurídica tem desencadeado uma revolução paradigmática. Nesse contexto, a Inteligência Artificial emerge como uma força motriz, promovendo mudanças significativas na maneira como os profissionais do direito abordam e solucionam questões complexas. Esta dissertação visa explorar de maneira abrangente a aplicação da Inteligência Artificial ao campo do direito, delineando suas nuances, seus tipos diversos e, adentrando na esfera do ChatGPT e na responsabilidade aplicada à IA. Inicialmente, se abordará o conceito fundamental de Inteligência Artificial, elencando suas diferentes categorias e apresentando uma análise aprofundada de como essas categorias se manifestam no âmbito jurídico. Em seguida, se explicará brevemente sobre o ChatGPT, um exemplo de IA generativa, capaz de interagir em linguagem natural, proporcionando uma visão tangível de como a tecnologia pode ser aplicada no cotidiano do advogado. Entretanto, não se pode ignorar os riscos inerentes ao avanço da Inteligência Artificial. Por isso, essa dissertação se propõe a examinar criticamente os potenciais desafios éticos e jurídicos associados ao uso dessa tecnologia no campo jurídico. A discussão sobre os riscos não é apenas uma precaução, mas uma reflexão necessária para garantir uma implementação responsável e benéfica dessas inovações. Ao explorar as aplicações práticas da Inteligência Artificial no cotidiano do advogado, se investigará como essa tecnologia pode otimizar processos, analisar grandes conjuntos de dados e, potencialmente, transformar a prestação de serviços jurídicos. Contudo, será destacada a necessidade premente de estabelecer diretrizes claras e responsabilidades ao incorporar a IA nas práticas jurídicas. Finalmente, a dissertação buscará delinear as responsabilidades éticas e legais que emergem do uso da Inteligência Artificial no domínio jurídico. Ao examinar casos específicos, padrões éticos e princípios legais pertinentes, propõe-se uma análise crítica sobre como a comunidade jurídica pode enfrentar os desafios éticos inerentes a essa evolução tecnológica. Assim, esta dissertação busca lançar luz sobre as complexidades e promessas da convergência entre Inteligência Artificial e Direito, fornecendo um arcabouço sólido para uma reflexão sobre o presente e o futuro dessa dinâmica relação.

Palavras-Chave: Inteligência artificial. Sistema. Responsabilidade. Inovação. Automação.

ABSTRACT

In the contemporary legal landscape, the growing intersection of technology and legal practice has sparked a paradigmatic revolution. In this context, Artificial Intelligence (AI) emerges as a driving force, bringing about significant changes in how legal professionals approach and address complex issues. This dissertation aims to comprehensively explore the application of Artificial Intelligence in the field of law, delineating its nuances, diverse types, and delving into the realm of ChatGPT and the responsibility applied to AI. Initially, we will address the fundamental concept of Artificial Intelligence, demystifying its different categories and providing an in-depth analysis of how these categories manifest in the legal domain. Subsequently, a brief explanation of ChatGPT, an example of generative AI capable of interacting in natural language, will be presented, offering a tangible insight into how the technology can be applied in the everyday life of a lawyer. However, the inherent risks associated with the advancement of Artificial Intelligence cannot be ignored. Therefore, this dissertation aims to critically examine the potential ethical and legal challenges associated with the use of this technology in the legal field. The discussion about risks serves not only as a precaution but as a necessary reflection to ensure a responsible and beneficial implementation of these innovations. While exploring the practical applications of Artificial Intelligence in the lawyer's daily routine, we will investigate how this technology can optimize processes, analyze extensive legal datasets, and potentially transform the delivery of legal services. However, the pressing need to establish clear guidelines and responsibilities when incorporating AI into legal practices will be highlighted. Finally, our dissertation will seek to outline the ethical and legal responsibilities that arise from the use of Artificial Intelligence in the legal domain. By examining specific cases, ethical standards, and relevant legal principles, we propose a critical analysis of how the legal community can address the inherent ethical challenges in this technological evolution. Thus, this paper aims to shed light on the complexities and promises of the convergence between Artificial Intelligence and Law, providing a robust framework for an informed reflection on the present and future of this dynamic relationship.

Keywords: Artificial intelligence. System. Responsibility. Innovation. Automation.

LISTA DE ABREVIATURA E SIGLAS

COTS	Produtos Comerciais Prontos
AIDA	Lei de Desenvolvimento de Inteligência Artificial
SVM	Máquinas de Vetores de Suporte
GANs	Redes Neurais Adversariais
RNNs	Redes Neurais Recorrentes
HMMs	Modelos Ocultos de Markov
ChatGPT	Transformadores pré-treinados generativos
PLN	Tecnologia de Processamento de Linguagem Natural
RPA	Automação de Processos Robóticos
MSI-NET	Comitê de especialistas em intermediários da Internet
ECHR	Convenção Europeia dos Direitos Humanos
DPIAs	Avaliações de Impacto de Proteção de Dados
IA	Inteligência Artificial
MIT	Instituto de Tecnologia de Massachusetts

SUMÁRIO

INTRODUÇÃO

O surgimento de novas tecnologias ao longo da história tem sido um catalisador para a evolução da sociedade. Desde a Revolução Industrial até a era digital, cada avanço tecnológico redefiniu a forma com que os seres humanos vivem e interagem. A eletricidade revolucionou a produção, a automação simplificou tarefas, e a internet conectou o mundo.

O impacto dessas inovações é vasto, permeando setores como saúde, educação, comunicação e economia. A tecnologia não apenas otimiza processos, mas também cria novas oportunidades. A medicina de precisão salva vidas, a educação online democratiza o conhecimento, e as redes sociais transformam a maneira como as pessoas se relacionam.

Com o direito, apesar de muitos ainda buscarem negar, as modificações também estão acontecendo, tanto pelo lado da aplicabilidade das novas ferramentas ao dia a dia dos advogados, quanto pela perspectiva de um conhecimento mais aprofundado das tecnologias, para, de alguma maneira, colaborar com a sociedade, encontrando maneiras de garantir que o uso seja benéfico e, ao mesmo tempo, responsável.

A inteligência artificial tem se tornado uma ferramenta valiosa no mundo jurídico, otimizando processos e oferecendo insights precisos. Na análise documental, por exemplo, algoritmos de IA podem acelerar a revisão de contratos, poupando tempo e reduzindo erros.

Além disso, sistemas de IA são utilizados para pesquisa jurídica, analisando vastas bases de dados em segundos e fornecendo informações relevantes para casos específicos. Isso permite que advogados acessem jurisprudência e doutrina de forma mais eficiente.

Já na predição de resultados judiciais, a inteligência artificial também desempenha um papel importante. Modelos preditivos podem analisar casos anteriores, identificando padrões e oferecendo boas sugestões sobre a probabilidade de decisões futuras.

Em resumo, a aplicação da inteligência artificial no mundo jurídico simplifica tarefas, acelera processos e aprimora a tomada de decisões, proporcionando uma prática jurídica mais eficiente e precisa.

Desta maneira, o direito, apesar de em muitos momentos ser visto à margem do ambiente tecnológico, possui papel extremamente relevante, tanto como usuário das ferramentas, o que naturalmente já impacta a sociedade, masa também como uma das partes principais durante o processo regulatório de sistemas que utilizam inteligência artificial. Como será melhor explicado adiante, o papel do direito na regulação da inteligência artificial é crucial para equilibrar o avanço tecnológico com a proteção dos direitos individuais e a preservação da ética. A IA, com seu potencial transformador, requer diretrizes que garantam uma aplicação responsável.

A regulação da IA abrange questões complexas, desde a transparência algorítmica até a responsabilidade por decisões autônomas. Direitos fundamentais, como privacidade e igualdade, necessitam ser preservados em um cenário onde algoritmos influenciam cada vez mais aspectos da vida cotidiana.

Além disso, as leis podem estabelecer padrões éticos para o desenvolvimento e uso da inteligência artificial. Ao definir limites e responsabilidades claras, o direito desempenha um papel preventivo contra potenciais abusos, assegurando que a tecnologia sirva ao bem comum.

Em síntese, a importância do direito na regulação da inteligência artificial reside em criar um ambiente que promova a inovação, ao mesmo tempo em que protege os indivíduos e a sociedade contra impactos negativos. Uma abordagem equilibrada é essencial para que a IA contribua positivamente para o progresso, sem comprometer valores fundamentais.

Ademais, a importância das novas tecnologias é evidente na capacidade de enfrentar desafios globais. Da inteligência artificial à energia sustentável, essas ferramentas oferecem soluções para questões prementes. Contudo, é crucial abordar questões éticas e garantir que o progresso tecnológico beneficie a humanidade como um todo.

Em síntese, o constante surgimento de novas tecnologias molda o futuro da humanidade. Seu impacto transcende o presente, influenciando gerações e definindo o curso da história. A sociedade, ao abraçar e guiar essas inovações, constroi um caminho para um amanhã mais promissor.

Por esse motivo, o presente trabalho tem como objetivo abordar o tema da inteligência artificial, demonstrando sua conexão com o ambiente jurídico. Para isso, a dissertação se dividirá em três capítulos, que de maneiras distintas demonstrarão

como o direito é parte importante desse processo de automação e de garantia de uso sustentável da inteligência artificial.

Sendo assim, o primeiro capítulo conceituará a inteligência artificial tendo como base os pensamentos de Scherer, diferenciará inteligência artificial discriminativa da generativa e abordará tanto explicações sobre o Chat GPT, quanto o Projeto de lei nº 2338/2023, que está em tramitação no Congresso Nacional. Nesse primeiro momento, é importante entender o contexto da IA, qual o patamar de desenvolvimento atual e como ela vem sendo utilizada pela sociedade, aprofundando e analisando cuidadosa e criteriosamente o Projeto de Lei, que, caso aprovar, regulará a utilização de sistemas que utilizem essa tecnologia.

Já o segundo capítulo, demonstrará que o universo jurídico tem totais condições de ser, além de protagonista no que tange a regulação e proteção dos dados e da segurança cibernética, um usuário da IA. Nessa seção, se aprofundará na maneira com que os advogados podem aplicar as novas tecnologias e quais vantagens podem ser auferidas nas rotinas, tanto dos escritórios de advocia, quanto dos departamentos jurídicos.

O terceiro e último capítulo explicará como o Comitê de Especialistas em Dimensões dos Direitos Humanos aplicadas ao Processamento Automatizado de Dados e Diferentes Formas de Inteligência Artificial entende a responsabilidade envolvida no uso dessas tecnologias. Por conseguinte, se abordará os principais dilemas enfrentados, os desafios encontrados quando da responsabilidade de uma empresa ou de um indivíduo, refletindo sobre a liberdade de expressão, o direito a privacidade e o devido processo legal.

1 APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL, BIG DATA E RPA AO DIREITO

Em alguns momentos, pode-se não perceber ou ser tão óbvio, mas vive-se atualmente na idade da inteligência artificial. A inteligência artificial está presente na vida das pessoas de diversas maneiras, algumas delas visivelmente perceptíveis, outras nem tanto, o que em determinados momentos pode parecer assustador.

Tarefas que há poucos anos eram performadas por profissionais altamente qualificados e com excelente formação, hoje podem ser facilmente cumpridas por sistemas computadorizados. Um claro exemplo dessa situação é trazido por Matthew U. Scherer, quando destaca o fato de uma inteligência artificial ter escrito um artigo para o renomado jornal francês *Le Monde* (Scherer, 2016, p. 354).

Em 2014, quando a nova tecnologia ainda não estava tão bem disseminada e de conhecimento público como nos dias de hoje, o também reconhecido periódico inglês, *The Guardian*, publicou o artigo do *Le Monde*: “The journalist who never sleep” (Eudes, 2014). Neste artigo, destacam-se algumas soluções tecnológicas capazes de redigir textos através da interpretação dos gostos dos leitores, bem como através da interpretação de números, como os provenientes de dados financeiros, por exemplo.

Entretanto, far-se-á relevante destacar a descrição de como um dos sistemas, o Quill, funciona. De acordo com um dos seus fundadores, Larry Birnbaum, professor da Northwestern University em Illinois, Estados Unidos: o Quill começa importando dados (tabelas, listas, gráficos) estruturados por outro software, depois outros sistemas inteligentes podem se encarregar de converter dados em diversos formatos (incluindo texto) em dados estruturados que podem ser utilizados por uma máquina. Desta forma, os escritores de robôs têm potencialmente acesso a todo o conhecimento humano. A próxima tarefa desempenhada pelo Quill é realizar a análise narrativa, quando os dados são classificados usando um método que se concentra exclusivamente na construção de uma narrativa, selecionando certos fatos, sublinhando ações, destacando números. A terceira e mais inovadora tarefa é gerar uma narrativa, ou seja, os algoritmos definem um plano, com uma lista de fatos e devido a um processo de modelagem, eles escolhem os ângulos editoriais adequados. Na prática, o resultado é uma mistura de palavras, linhas de código,

gráficos – uma representação que só as máquinas podem entender (Birnbaum, 2016, p. 2).

Através da descrição feita por Birnbaum pode-se averiguar que, mesmo em 2014, já existiam tecnologias capazes de capturar dados, mesmo que complexos, interpretá-los e transformá-los em um texto de altíssima qualidade. Ao mesmo tempo, a captura de dados e tendências pessoais demonstradas na internet já era de extrema relevância, ao ponto de viabilizar que artigos personalizados observando os gostos de determinados leitores fossem criados.

Uma preocupação trazida pelo artigo de Birnbaum era a substituição de jornalistas pela inteligência artificial, uma preocupação legítima e bastante comum, mas que não se concretizou e não se concretizará, visto que a inteligência artificial não substituirá o ser humano nas tarefas legitimamente humanas e que necessitam de posicionamento e subjetividade. De acordo com o próprio artigo publicado no periódico, a intenção é que mais artigos sejam escritos e com mais foco nas particularidades do público, aumentando o engajamento e o alcance dos textos, mas de forma alguma, os principais temas seriam exclusivamente cobertos por inteligência artificial.

Scherer destaca que grandes companhias, como Google, Facebook e Amazon estão cada vez mais investindo nessa tecnologia, seja através de pesquisas, construindo laboratórios ou apoiando start-ups especializadas (Scherer, 2016, p. 355). Por conta disso, a IA vem a cada dia atravessando fronteiras e adentrando novos mercados, tendo novas funcionalidades e tornando-se mais potente, já que quanto maior a quantidade de dados disponíveis e tratados, maiores a condições desta tecnologia obter resultado mais acurados.

O potencial desses novos avanços trouxe consigo preocupações em muitos setores, que sugeriram inclusive uma regulação do governo sob a IA. Um desses setores reticentes, curiosamente foi o de *tech*, que defendeu que a tecnologia poderia aumentar o desemprego e teve medo de que novas tecnologias caíssem em desuso (Scherer, 2016, p. 355).

Ter os líderes de empresas de tecnologia receosos com a inteligência artificial e buscando maneiras de regulá-la somente demonstra como há egoísmo na sociedade mundial. O desenvolvimento da tecnologia, enquanto os beneficiava, era perfeito e o desemprego que suas próprias criações traziam não era enxergado e não preocupava, entretando, quando a possibilidade de seus próprios negócios

serem impactados surge, o cenário se altera, podendo ser traçado um paralelo perfeito com o poema “Intertexto”[1] de Bertold Brecht (Frez, 2017).

Elon Musk dono da rede social Twitter, durante uma entrevista na MIT’2014 *AeroAstro Centennial Symposium* chegou a defender a regulação da inteligência artificial, a nível nacional e internacional, já que para ele essa tecnologia seria a maior ameaça existencial da raça humana (Graef, 2014).

Com a maior disseminação da IA, é esperado que alguns questionamentos e dilemas relacionados ao direito passem a ser enfrentados, principalmente no que tange a responsabilidade civil e/ou criminal. Scherer cita o exemplo dos carros dirigidos sem motoristas, já utilizados nos Estados Unidos e que foram postos em circulação sem que houvesse lei regulando a sua possibilidade ou jurisprudência que pudesse ser aplicada (Scherer, 2016, p. 356).

Somado a isso, não parece haver uma grande quantidade de faculdades de direito discutindo a regulação da Inteligência Artificial, fazendo pesquisas e sendo capaz de colaborar com o governo no que tange a como enfrentar os dilemas que o futuro trará. É bem verdade que, apesar do direito ter o dever de caminhar ao lado da sociedade, modernizando-se em conjunto, essa não vem sendo uma realidade nos últimos anos. O que se tem visto é um direito que muitas vezes é deficiente e que se mostra conservador, investindo e se interessando pouquíssimo em pesquisas sobre novas tecnologias, por exemplo.

De certa forma, não surpreende que o assunto da regulamentação da IA tenha sido recebido com silêncio pelo mundo do direito. Os métodos tradicionais de regulamentação - como o controle de licenciamento de produto, supervisão de pesquisa e desenvolvimento e responsabilidade civil -, parecem particularmente inadequados para gerenciar os riscos associados a máquinas inteligentes e autônomas.

Segundo Scherer, a regulação é algo difícil porque a pesquisa e o desenvolvimento de IA podem ser discretos (exigindo pouco infraestrutura física), e diferentes componentes de um sistema de IA podem ser projetados sem coordenação consciente. Além disso, a IA pode ser difusa, ou seja, dezenas de

[1] “Primeiro levaram os negros/ Mas não me importei com isso/Eu não era negro/ Em seguida levaram alguns operários/ Mas não me importei com isso/ Eu também não era operário/ Depois prenderam os miseráveis/ Mas não me importei com isso/ Porque eu não sou miserável/ Depois agarraram uns desempregados/ Mas como tenho meu emprego/Também não me importei/ Agora estão me levando/ Mas já é tarde./ Como eu não me importei com ninguém/ Ninguém se importa comigo”

indivíduos em localizações geográficas dispersas podem participar em um projeto de IA e observadores externos podem não ser capazes de detectar recursos potencialmente prejudiciais de um sistema de IA (Scherer, 2016, p. 356).

A natureza da IA cria questões de previsibilidade e controle que podem tornar a regulamentação ineficaz, especialmente se um sistema de IA representar um risco catastrófico. O fato da IA poder se desenvolver cada vez mais e alcançar profundidades inimagináveis também dificulta a regulação, já que é algo em constante modificação. Além disso, a regulamentação em qualquer estágio é complicada pela dificuldade em definir o que, exatamente, “inteligência artificial” significa.

De acordo com Scherer, o crescimento da IA na economia e na sociedade traz consigo desafios práticos e conceituais para o sistema legal. Muitos dos desafios práticos são provenientes da maneira como a IA é pesquisada e desenvolvida e dos básicos problemas de controlar as ações de máquinas autônomas. Já os desafios conceituais vêm da dificuldade de estabelecer responsabilidade moral e legal para danos causados por essa tecnologia e do quebra-cabeças para definir exatamente o que a inteligência artificial significa.

Alguns desses problemas são exclusivos da inteligência artificial, mas outros podem ser também encontrados quando se trata da demais tecnologias. Por esse motivo, Scherer sugere que o sistema jurídico terá grandes dificuldades para gerenciar o aumento da IA e garantir que as partes prejudicadas recebam compensação quando um sistema como esse causar danos (Scherer, 2016, p. 358).

1.1 O que é inteligência artificial?

Para que seja possível regular algo, o primeiro passo é sempre compreender o que se está interpretando e regulando, entretanto, nesse caso, essa tarefa não é uma tarefa fácil, já que não há uma definição clara e objetiva da inteligência artificial, nem mesmo entre os especialistas.

De acordo com Scherer, a dificuldade em definir o que seria inteligência artificial não mora no entendimento do que seria artificial, mas sim na compreensão da “inteligência”. Isso ocorre porque historicamente, entendeu-se que os seres humanos são os únicos a possuir inteligência e esse seria o principal diferencial dessa raça para as demais habitantes do planeta Terra (Scherer, 2016, p. 359).

John McCarthy, pioneiro na definição de inteligência artificial, defendeu que “não há definição sólida de inteligência que não dependa do relacionamento com a inteligência humana”, porque “não é possível definir quais procedimentos tecnológicos desejamos chamar de inteligência” (McCarthy, 2007, p. 2-3). As definições de inteligência, portanto, variam amplamente e se concentram em características humanas interconectadas e essas características incluem consciência, uso da linguagem, capacidade de aprender, de abstrair, se adaptar e raciocinar (Adelson, 2005, p. 22; Scherer, 2016, p. 360).

Desta maneira, como se pode perceber, a elucidação do que seria inteligência é algo que ultrapassa o entendimento do que seria inteligência artificial. Quais seriam os elementos essenciais para caracterizar a inteligência? Poderia algo que não é humano ser auto-consciente, raciocinar, abstrair e aprender? E o que seria inteligência no sentido geral?

Stuart Russel e Peter Norvig apresentam, de maneira brilhante, oito diferentes definições do que seria inteligência, divididas em quatro principais grupos, que seriam: “pensar humanamente, agir humanamente, pensar racionalmente e agir racionalmente” (Russel; Norving, 2010, p. 2). Considerando o estudo dos autores, o que mais se tem feito é definir inteligência artificial focando no conceito de máquinas que trabalham para atingir um objetivo, agindo, portanto, de maneira racional. Todavia, a compreensão do que seria o termo “objetivo” também não é clara, fazendo com que a subjetividade da IA seja ainda maior.

Para que se possa tratar da regulação da IA, existe a necessidade de se apresentar uma definição clara. Mesmo que se observe que essa tecnologia sob a perspectiva do “agir racionalmente”, ainda seria uma descrição vazia, visto que existem outras tecnologias que agem racionalmente e não por isso são consideradas inteligência artificial. Além disso, não seria possível garantir que, por agirem racionalmente as inteligências artificiais não ofereceriam um risco para a sociedade (Scherer, 2016, p. 362).

Sendo assim, Scherer culmina por entender inteligência artificial como sendo “máquinas capazes de performar tarefas, que se performadas por humanos, necessitariam do uso da inteligência” (Scherer, 2016, p. 362).

Algumas características da inteligência artificial merecem destaque, uma vez que são capazes de atestar a sua diferenciação das demais tecnologias surgidas anteriormente. Uma dessas características é a autonomia, ou seja, os sistemas que

utilizam inteligência artificial não precisam de uma supervisão ou de um controle direto de um humano, podendo performar suas atividades de maneira automática.

Atualmente, diversos exemplos dessa autonomia podem ser vistos no dia a dia da sociedade mundial e a tendência é que em alguns anos a presença seja ainda maior, através de carros dirigidos sem a presença de um condutor humano, por exemplo. Em conjunto com esse desenvolvimento, certamente, novas problemáticas serão (e já são) geradas e, consequentemente, necessitarão de um apoio legal, principalmente no que tange o direito trabalhista, responsabilidade civil e propriedade intelectual, por exemplo.

Outra característica marcante da inteligência artificial, que a faz diferenciar das demais tecnologias disponíveis e também dos seres humanos, é a previsibilidade. A IA é capaz de captar, interpretar e utilizar uma grande gama de dados, o que proporciona que ela avalie uma também gigantesca quantidade de possibilidades antes de tomar uma ação e, por isso, em alguns momentos, as escolhas realizadas pelo sistema provido de IA será diferente da que normalmente um humano tomaria.

Os humanos, por não possuirem a mesma capacidade de armazenamento e interpretação, e também por terem emoções envolvidas, acabam por considerar algumas dessas soluções trazidas pela IA, como inovadoras e criativas, quando na verdade, elas nada mais são do que a tecnologia sendo utilizada de maneira plena (Scherer, 2016, p. 363). Todavia, apesar dessa criatividade parecer brilhante, em alguns momentos ela pode se configurar como um risco, uma vez que soluções unicamente baseadas em dados e em cálculos matemáticos podem gerar ações desequilibradas e sem consonância com a moral e com as boas práticas de preservação da raça humana.

Há ainda um ponto de extrema relevância que merece destaque, que seria o controle sobre os sistemas regidos pela inteligência artificial. Em grande parte das vezes, essas tecnologias são programadas para possuir autonomia e agir de maneira independente, visto que é essa liberdade e permissão para criar que fazem a IA ser tão diferente e revolucionária.

Todavia, em alguns momentos, é possível que os humanos tenham alguma (ou muita) dificuldade para ter controle sobre os sistemas. De acordo com Scherer (2016, p. 366), existe uma vasta quantidade de exemplos de situações onde o controle pode ser perdido, dentre elas: arquivo corrompido, quebra de segurança e

falha na programação. E é nesse momento que surge uma das maiores preocupações relacionadas a IA.

Possuir autonomia e ser capaz de apresentar caracteristicas similares às da inteligência humana, tais como aprendizado e adaptação dificultam o controle sob as ações tomadas pela tecnologia e gera, consequentemente, riscos imensuráveis.

Scherer (2016, p. 367) divide a perda de controle em duas diferentes categorias: a perda de controle local, que ocorre quando os criadores/responsáveis pela operação do sistema não são mais capazes de controlá-lo e; a perda de controle geral, que acontece quando nenhum ser humano consegue mais controlar o sistema. Obviamente, o surgimento de um sistema como esses ocorre a partir da programação de um objetivo a ser atingido pela inteligência artificial, de modo que todas as ações tomadas autonomamente por ela deverão seguir na direção do que foi programado por um ser humano como sendo o alvo do desenvolvimento.

O fato de ser o ser humano a programar essa inteligência e ditar quais seriam os objetivos finais gera uma segurança e uma esperança de que os riscos sejam mitigáveis, entretanto, alguns especialistas no tema sugerem que, em alguns momentos, o sistema pode estar programado para perseguir um alvo e, mesmo que os resultados obtidos no caminho não estejam de acordo com o que os seres humanos responsáveis por sua operação e programação imaginavam, o objetivo permanece sendo perseguido.

Desta maneira, é facilmente perceptível que um dos principais pontos de atenção quando se está estudando e/ou aplicando a inteligência artificial é observar com atenção os objetivos programados inicialmente para os sistemas, levando sempre em consideração que, por mais que se trate de uma inteligência capaz de realizar muitas tarefas antes exclusivas dos seres humanos, ainda não há a interpretação, a racionalidade e a moral que aqueles possuem. Sendo assim, nem algumas ocasiões, pode ser que o objetivo seja fielmente perseguido mesmo que para isso tenha que atentar contra uma vida, por exemplo.

Nesse sentido, de acordo com Scherer (2016, p. 368), existem ainda alguns especialistas mais conservadores que entendem que a inteligência artificial pode resistir inclusive a todos os esforços humanos para controlar suas ações, o que significaria que se estaria em face de um risco contra a humanidade. Mas por que isso aconteceria?

Como dito anteriormente, esses sistemas possuem grande autonomia e, em

dado momento, eles podem melhorar o seu próprio hardware a ponto de possuir maior capacidade do que a própria raça humana. Russel e Norvig (2010, p. 1037) acreditam, inclusive, que não é uma tarefa fácil garantir que os objetivos dos sistemas permanecem os mesmos daqueles que o construiram e prepararam seu design.

Não é preciso concordar com a plausibilidade de cenários de risco existencial para reconhecer que surgirão problemas de controle e supervisão à medida que os sistemas de IA se tornarem mais poderosos, sofisticados e autônomos. Hoje, os sistemas de IA já têm a capacidade de executar alguns comandos autonomamente, como por exemplo, transações de ações em uma escala de tempo medida em nanossegundos, tirando dos humanos a capacidade de intervir em tempo real. O "flash crash" de 2010 demonstrou que a interação entre sistemas de negociação algorítmica pode ter um impacto econômico enorme em um período de tempo curto. Felizmente, pelo menos teoricamente, os resultados dessas transações de ações podem ser revertidos para a maioria dos investidores humanos.

O ponto nevráugico do estudo e da regulação da inteligência artificial talvez esteja exatamente na possibilidade de revertê-la, uma vez que dessa maneira, caso se vislumbre qualquer indício de que o objetivo inicialmente programado não está sendo atingido e que o desenvolvimento autônomo está ocorrendo de maneira a fugir ao controle não só do seu programador, mas também de todo e qualquer programador, existe a opção de interrromper a atividade.

De acordo com Scherer, far-se-á necessário aprofundar na pesquisa e no desenvolvimento, levando em conta alguns de seus aspectos, quais sejam: discreção, difusão, distinção, e o fato de ser opaca. Do ponto de vista regulatório, os pontos mais problemáticos da IA não estão relacionados à própria tecnologia, mas sim à forma como a pesquisa e desenvolvimento em IA são conduzidos.

A falta de visibilidade, ou seja, a discreção, se refere ao fato de que o trabalho de desenvolvimento da IA pode ser realizado sem uma infraestrutura claramente visível. Já a difusão/dispersão significa que as pessoas trabalhando em diferentes componentes de um sistema de IA podem estar localizadas distantes umas das outras. Além disso, diferentes componentes de um sistema de IA podem ser projetados em lugares e tempos diferentes, sem uma coordenação consciente. Por fim, a opacidade indica a possibilidade de que o funcionamento interno de um sistema de IA possa ser mantido em segredo e não seja suscetível a engenharia

reversa. Todas essas características são compartilhadas, em diferentes graus, por trabalhos de pesquisa e desenvolvimento em várias tecnologias da Era da Informação, mas apresentam desafios particularmente únicos no contexto da IA (Scherer, 2016, p. 369).

As fontes de risco público que caracterizaram o século XX, como tecnologia nuclear, bens de consumo produzidos em massa, poluição industrial em grande escala e produção de grandes quantidades de substâncias tóxicas, exigiram investimentos significativos em infraestrutura. Isso simplificou o processo regulatório. O alto custo de construção das instalações necessárias, aquisição dos equipamentos necessários e contratação da mão de obra significava que apenas grandes corporações eram capazes de gerar a maioria das fontes de risco público fora do âmbito governamental. Além disso, as pessoas responsáveis pela instalação, operação e manutenção da infraestrutura geralmente precisavam estar presentes fisicamente no local onde a infraestrutura estava localizada. A visibilidade física da infraestrutura e das pessoas envolvidas tornava altamente improvável que os riscos públicos pudessem ser gerados de forma clandestina. Dessa forma, os reguladores encontravam pouca dificuldade em determinar quem e onde poderiam surgir potenciais fontes de risco público (Scherer, 2016, p. 369).

Ao contrário do que ocorreu com no século XX, a pesquisa e desenvolvimento em IA podem ser realizados de forma relativamente discreta, o que é uma característica compartilhada por muitas outras tecnologias da Era da Informação e acaba por dificultar a regulação e o controle.

Em 2009, o Professor John McGinnis (2010, p. 15) afirmou que "a pesquisa em inteligência artificial é realizada por instituições que não são mais ricas do que faculdades e talvez até exijam recursos menos substanciais". Essa declaração, na verdade, superestimou os recursos necessários para participar do desenvolvimento de IA, especialmente com o aumento da programação de código aberto.

Em termos simples, uma pessoa não precisa dos recursos e das instalações de uma grande corporação para escrever um código de computador. Qualquer pessoa que possua um computador ou até mesmo um smartphone e uma conexão com a Internet pode contribuir para projetos relacionados à IA. Os indivíduos agora têm a capacidade de participar do desenvolvimento da IA de qualquer lugar e essa possibilidade é o que diferencia mais a IA das fontes anteriores de risco público.

Desta forma, os desenvolvedores de um software que possua inteligência

artificial não precisam fazer parte da mesma organização, ou até mesmo de qualquer organização. Uma prova clara disso é que já existem várias bibliotecas de inteligência artificial com códigos abertos, permitindo que indivíduos espalhados façam dezenas de modificações nessas bibliotecas diariamente. Essas modificações podem até ser feitas de forma anônima, no sentido de que a identidade física das pessoas que as realizam não é facilmente encontrada.

Considerando o pensamento de Scherer (2016, p. 370), o próprio programa de IA pode ter componentes de software retirados de várias dessas bibliotecas, cada uma delas sendo construída e desenvolvida independentemente das outras. Um indivíduo que participe da construção de uma biblioteca de código aberto muitas vezes não tem conhecimento prévio de quais outros indivíduos ou entidades podem usar essa biblioteca no futuro. Componentes retirados dessas bibliotecas podem então ser incorporados à programação de um sistema de IA que está sendo desenvolvido por uma entidade que não participou da criação da biblioteca.

Essas características não estão restritas apenas a projetos de código aberto ou materiais disponíveis gratuitamente, visto que muitos sistemas de computador modernos utilizam componentes comerciais prontos para uso ou *Commercial off-the-shelf* – COTS (Dewar, 2010). A facilidade com que esses componentes podem ser adquiridos incentiva a sua maximização para controlar os custos, mesmo com os possíveis problemas de segurança associados ao uso de componentes de software desenvolvidos completamente fora do controle dos desenvolvedores do sistema (Woody; Ellison, 2013; Scherer, 2016, p. 371).

A programação de IA moderna segue essa tendência, ou seja, poucos sistemas de IA são construídos do zero, usando componentes e códigos totalmente criados pelos próprios desenvolvedores de IA. Além disso, é provável que os componentes físicos de um sistema de IA sejam fabricados por entidades distintas daquelas que desenvolveram a programação do sistema de IA (Scherer, 2016, p. 371).

Embora seja comum encontrar componentes desenvolvidos separadamente em máquinas complexas, a discreção e a interação entre componentes de software e hardware em sistemas de computador modernos já rivalizam ou até superam as tecnologias anteriores, e essa complexidade tende a aumentar ainda mais com o avanço da IA (Scherer, 2016, p. 371).

Alguns sistemas de IA serão construídos principalmente com componentes

COTS ou de código aberto, enquanto outros utilizarão principalmente componentes de programação e físicos projetados e desenvolvidos especificamente para o projeto de IA em questão. No entanto, devido às vantagens de custo, é quase certo que alguns sistemas de IA funcionarão com uma combinação de componentes de hardware e software provenientes de várias empresas. A interação entre inúmeros componentes e a diversidade de localizações geográficas das empresas envolvidas tornará muito mais complexa a implementação de um regime para gerenciar os riscos associados à IA (Scherer, 2016, p. 371).

Sendo assim, ao passo que se aprofunda na compreensão acerca das particularidades da IA, percebe-se o quão complexa é a sua regulação e o seu controle pelos seres humanos. Em certos momentos, pode parecer um tanto quanto assustador pensar que indivíduos em países diversos podem contribuir para a construção de um sistema de IA, mesmo sem ter conhecimento disso, o que dificulta, e muito, a mitigação ddos riscos contra a humanidade.

Esse risco cresce não somente porque pessoas diferentes estariam ajudando sem nem possuirem conhecimento, mas também por, em alguns momentos, poder haver perda da noção acerca dos objetivos da criação dos códigos e, como foi visto anteriormente, esses objetivos são exatamente a chave para a manutenção do controle sobre a IA. Caso os objetivos não estejam sendo atingidos de acordo com o seu propósito inicial, e caso a autonomia esteja sendo considerada prejudicial, o melhor a ser feito seria resetar a funcionalidade, o que talvez não seja tão simples quando se trata de um código aberto, criado por uma pessoa anônima, por exemplo.

Nesse momento, paira uma grande preocupação acerca de quem poderia solucionar a questão, já que nem os próprios programadores parecem ser capazes de, a longo prazo, controlar os riscos trazidos pela IA. Seria o direito o melhor caminho?

De acordo com Scherer (2016, p. 374), apesar das características problemáticas da IA, há motivos válidos para acreditar que mecanismos legais poderiam ser utilizados para diminuir os riscos públicos que a IA apresenta sem inibir a inovação. Muitos dos problemas identificados anteriormente são simplesmente lacunas na legislação atual, e essas falhas podem ser solucionadas de diversas maneiras.

Elaborar uma definição funcional de IA certamente é desafiador, porém estabelecer definições jurídicas precisas para termos imprecisos não é um desafio

exclusivo da IA. Qualquer definição legal com fins de responsabilidade ou regulamentação provavelmente será abrangente demais ou restritiva demais, mas isso também não é um problema desconhecido para o sistema jurídico enfrentar (Scherer, 2016, p. 374).

Da mesma forma, as questões relacionadas à previsibilidade e causalidade precisam ser abordadas, mas os tribunais sempre precisaram ajustar as regras de causalidade próxima à medida que a tecnologia mudou e se desenvolveu. O problema do controle apresenta desafios consideráveis em termos de limitar os danos causados pelos sistemas de IA após seu desenvolvimento, mas isso não torna mais difícil regular ou orientar o desenvolvimento da IA antecipadamente (Scherer, 2016, p. 374).

É possível facilmente perceber que em diversos momentos ao longo da história, o direito absorveu para si a obrigação de criar definições mais objetivas a fim de regular assuntos que, a princípio também pareciam sombrios. A cada novidade que surge e que, consequentemente, necessita de uma maior atenção através da regulação, o direito assume esse papel. Sendo assim, com a inteligência artificial não deveria ser diferente.

Como se sabe, não é a primeira vez que o direito se depara com alguma situação onde existe falta de transparência e, por isso, a lei já disponibiliza mecanismos para enfrentar esse tipo de situação. A discrepância da IA também é compartilhada por muitas outras tecnologias modernas e até mesmo antigas.

Um exemplo trazido por Scherer (2016, p. 374) são os automóveis, que há muito tempo são fabricados utilizando componentes de várias empresas, e os tribunais desenvolveram regras para atribuir responsabilidade quando ocorrem danos causados por defeitos em vários desses componentes. A falta de transparência poderia ser reduzida tanto por meio de legislação direta, exigindo a publicação do código e das especificações dos sistemas de IA disponibilizados para venda comercial, como de forma indireta, através de incentivos fiscais ou padrões de responsabilidade civil que limitem a responsabilidade das empresas que tornam seus sistemas de IA mais transparentes.

Os desafios apresentados pela natureza potencialmente dispersa e discreta da pesquisa e desenvolvimento de IA parecem um pouco mais difíceis de resolver à primeira vista. No entanto, o simples fato de que a IA possa ser desenvolvida de maneira dispersa e discreta não significa que o progresso da IA seguirá um caminho

radicalmente diferente das fontes anteriores de riscos públicos.

De acordo com Scherer, já existem indícios na indústria de que o desenvolvimento da IA, assim como a maioria das tecnologias do século XX, será em grande parte impulsionado por entidades comerciais e governamentais, em vez de pequenos atores privados, o que pode faciliar bastante a sua regulação e o seu controle. O potencial comercial dessa tecnologia já desencadeou uma verdadeira corrida pela IA, e um exemplo disso é o grande investimento feito pela Google e outras grandes empresas, que têm projetos significativos de IA. O epicentro e a solução da pesquisa e desenvolvimento pode, portanto, estar no mesmo lugar dos riscos públicos do século XX - grandes corporações altamente visíveis (Scherer, 2016, p. 375).

Nesse sentido, a partir do momento que se tem as grandes empresas e até mesmo o governo investindo nesse tipo de tecnologia, um pouco mais de segurança passa a surgir, visto que essas empresas por possuirem o tamanho que possuem, necessitam implementar um volume grande de controles, bem como apresentar periodidamente seus resultados, reportando para seus investidores e, às vezes, para a população em geral as evoluções, os problemas e os riscos enfrentados. Por isso, a falta de transparência quando se está diante de uma empresa como a Google, por exemplo, pode ser um pouco mitigada, principalmente no que tange a responsabilização e o controle dos objetivos do software.

No entanto, de acordo com Scherer (2016, p. 376), o aumento do investimento do setor privado em IA restringe as opções disponíveis para o governo. O alto investimento do setor privado no desenvolvimento de IA torna improvável que pesquisas subsidiadas pelo governo sobre segurança da IA tenham um impacto significativo se forem realizadas isoladamente.

A menos que haja um gasto público extraordinariamente alto, o investimento do governo em pesquisa de IA será insignificante em comparação com o investimento do setor privado. A não ser que ocorra um evento de grande escala, como a Segunda Guerra Mundial, é improvável que haja um apoio público para um grande investimento governamental em projetos de IA (Scherer, 2016, p. 376).

Além disso, se o objetivo do investimento público em IA for apenas pesquisar a segurança da IA e divulgar informações sobre como desenvolver IA segura, ainda será necessário algum tipo de mecanismo para incentivar ou exigir que os desenvolvedores incorporem recursos de segurança em seus sistemas. Portanto,

embora a pesquisa subsidiada pelo governo possa complementar um quadro legal mais amplo para a IA, não seria uma suficiente para os riscos públicos gerados (Scherer, 2016, p. 376).

Felizmente, os organismos legais e regulatórias do mundo industrializado oferecem uma ampla gama de ferramentas que podem influenciar o desenvolvimento e a operação da IA de maneiras social e economicamente benéficas. Mesmo que um aspecto da IA não possa ser facilmente regulado diretamente por uma agência administrativa, ele pode responder aos incentivos indiretos fornecidos pela responsabilidade civil. O poder legislativo, agências reguladoras e tribunais oferecem mecanismos que podem ajudar a orientar o desenvolvimento da IA de forma positiva (Scherer, 2016, p. 376).

Scherer, no decorrer do estudo dele analisa as habilidades de três importantes instituições, quais sejam: poder legislativo nacional, agências reguladoras e sistema common law de responsabilidade civil - especialmente no que diz respeito à gestão dos riscos públicos apresentados pela IA (Scherer, 2016, p. 377).

Os procedimentos legais de todas as três instituições possuem características semelhantes. Os recursos financeiros e profissionais disponíveis para grandes empresas e pessoas ricas concedem a elas uma maior capacidade de influenciar as tomadas de decisão em todos os contextos institucionais. Essa tendência se reflete através do lobby realizado por grupos de interesse concentrados no poder legislativo e agências reguladoras (Scherer, 2016, p. 377).

No sistema de responsabilidade civil, o acesso a recursos financeiros maiores proporciona aos litigantes a capacidade de gastar mais dinheiro em investigação, advogados e especialistas. Acadêmicos têm debatido incessantemente sobre qual instituição é mais prejudicada por essas disparidades (Scherer, 2016, p. 377).

Por enquanto, é suficiente observar que o acesso a recursos financeiros maiores geralmente se traduz em uma capacidade superior de influenciar políticas nos três contextos. Outras características comuns às três instituições, e que não afetam claramente nenhuma delas mais do que as outras, incluem restrições de recursos e a possibilidade de corrupção por parte dos principais tomadores de decisão (Scherer, 2016, p. 377).

Além dessas características comuns, nenhuma instituição detém o monopólio de uma determinada competência. Por exemplo, embora as agências reguladoras

geralmente possuirem vantagem sobre tribunais e poderes legislativos em termos de conhecimento especializado em determinadas áreas, tribunais e poderes legislativos podem reduzir essa diferença consultando especialistas próprios. E embora poderes legislativos e agências tenham mais liberdade do que tribunais para agir de forma preventiva e tomar medidas para evitar danos, é razoável questionar com que frequência eles exercem essa liberdade na prática (Scherer, 2016, p. 378).

Os estudiosos têm, em grande parte, negligenciado o papel regulatório do poder legislativo, optando em vez disso por se concentrarem nos processos judiciais e administrativos. Essa omissão parece um tanto intrigante considerando-se que se está no século XXI e é crescente a importância da intervenção legislativa direta como forma de controle social e econômico. No século XX, ocorreu a criação de códigos tributários complexos e a gradual substituição dos crimes do direito comum por códigos penais. Esquemas estatutários também cada vez mais substituíram o direito comum na definição das regras substanciais que regem a falência, o trabalho, a saúde pública, a propriedade imobiliária e o transporte pessoal (Scherer, 2016, p. 378).

Apesar da relativa escassez de estudos discutindo as forças e fraquezas institucionais dos poderes legislativos, é possível perceber algumas características gerais: (1) legitimidade democrática; (2) uma relativa falta de especialização; e (3) a capacidade de delegação. Essas características tornam os poderes legislativos o corpo ideal para estabelecer o ponto de partida de um esquema regulatório e estabelecer os princípios fundamentais que orientam o desenvolvimento de políticas, embora não sejam adequadas para tomar decisões sobre o conteúdo específico das regulamentações (Scherer, 2016, p. 378).

As leis aprovadas por órgãos legislativos compostos por representantes eleitos têm uma responsabilidade mais forte em refletir a vontade popular do que as regras administrativas ou doutrinas. Isso confere às promulgações legislativas uma legitimidade democrática maior do que as regras das agências ou decisões dos tribunais. Essa vantagem decorre tanto do fato de os deputados e senadores serem escolhidos por meio de eleições regulares quanto da maior abertura do legislativo ao contato direto com o público em geral. Portanto, a população tende a preferir que o legislativo tome as decisões em questões de políticas sociais fundamentais (Scherer, 2016, p. 378).

No entanto, embora o processo democrático forneça aos legislativos sua

principal reivindicação de preeminência na formulação de políticas, os eleitores geralmente votam com base no conjunto completo de visões de um candidato, em vez de em uma única questão, e raramente conhecem os detalhes exatos de qualquer projeto de lei específico no momento em que entram na urna, minando assim o princípio idealista de que a ação legislativa é uma expressão da vontade popular.

A necessidade de ser reeleito e o custo das campanhas legislativas também limitam a capacidade dos legisladores de fazer julgamentos informados sobre qualquer projeto de lei específico. Os legisladores dedicam um tempo considerável à campanha e à captação de recursos, reduzindo a quantidade de tempo que passam com os assuntos legislativos e em contato amplo com os eleitores. A pressão de grupos de interesse importantes pode levar um legislador a apoiar políticas que seus eleitores se opõem e a se opor a políticas que eles apoiam (Scherer, 2016, p. 379).

Apesar dessas preocupações, os poderes legislativos continuam sendo as instituições mais adequadas para tomar decisões políticas que envolvem valores. Os funcionários das agências são nomeados em vez de eleitos; os juízes devem seguir a lei, mesmo quando ela se desvia da vontade popular; e abandonar esses princípios para tornar as agências e os tribunais mais responsivos democraticamente minaria as forças exclusivas dessas instituições. Por padrão, então, os legisladores estão mais bem preparados para tomar decisões sobre questões em que a legitimidade democrática é uma prioridade (Scherer, 2016, p. 379).

Qualquer regime regulatório de IA deve ter a aprovação pública que vem com a aprovação legislativa. A ponderação de valores é inerente tanto para determinar o nível de risco público aceitável quanto para decidir se existem determinadas áreas (por exemplo, funções militares e policiais) em que os tomadores de decisão humanos nunca devem transferir a responsabilidade para máquinas autônomas. Para garantir que instituições com forte legitimidade democrática tomem essas decisões, os poderes legislativos devem estabelecer o ponto de partida para a regulamentação de IA especificando os objetivos e propósitos de qualquer regime regulatório de IA (Scherer, 2016, p. 379).

Uma fraqueza crítica dos poderes legislativos em relação à regulamentação de novas tecnologias é a falta de conhecimento especializado. As agências normalmente contam com especialistas que possuem conhecimento especializado

no campo em questão e os tribunais podem recorrer a testemunhas especializadas para obter o conhecimento técnico necessário para decidir um caso específico. Por outro lado, os legisladores geralmente precisam de audiências de comitês e do contato com grupos de lobby para ter acesso a opiniões especializadas relevantes sobre a legislação proposta (Scherer, 2016, p. 380).

As desvantagens de depender de lobistas para obter informações técnicas são óbvias, e a utilidade das audiências de comitê é questionável. Apenas o pequeno subconjunto da legislatura que faz parte do comitê ouvirá o testemunho dos especialistas, e mesmo esses legisladores não podem se dar ao luxo de gastar uma quantidade excessiva de tempo conduzindo audiências sobre qualquer assunto ou projeto específico. Além disso, especialmente no Congresso dos Estados Unidos, a influência dos comitês legislativos diminuiu perceptivelmente nos últimos anos. Isso limita a capacidade de uma legislatura de tomar decisões políticas informadas para tecnologias emergentes, onde uma compreensão adequada dos aspectos relevantes de uma tecnologia pode depender do acesso a conhecimento técnico (Scherer, 2016, p. 380).

Scherer (2016, p. 393), ao final do seu artigo, apresenta uma proposta de regime regulatório para IA. De acordo com ele, o objetivo dessa proposta não é fornecer um plano completo para um regime regulatório de IA, mas sim iniciar uma discussão sobre a melhor forma de gerenciar os riscos públicos associados à IA sem restringir a inovação.

Com esse intuito, propõe uma legislação, a Lei de Desenvolvimento de Inteligência Artificial (AIDA), que criaria uma agência responsável por certificar a segurança de sistemas de IA. Em vez de conceder à nova agência poderes semelhantes aos da *Federal Drug Administration* dos Estados Unidos, para proibir produtos considerados inseguros, a AIDA estabeleceria um sistema de responsabilidade em que os projetistas, fabricantes e vendedores de programas de IA certificados pela agência estariam sujeitos a uma responsabilidade civil limitada, enquanto programas não certificados oferecidos para uso ou venda comercial estariam sujeitos a uma responsabilidade conjunta e solidária rigorosa.

A AIDA aproveitaria as respectivas forças institucionais de poderes legislativos, agências e tribunais, e levaria em consideração os aspectos singulares da pesquisa em IA, que tornam especialmente desafiadora a sua regulamentação. Ela se beneficia da legitimidade democrática dos poderes legislativos, atribuindo aos

legisladores a tarefa de estabelecer os objetivos e propósitos que orientam a regulamentação da IA e delega a tarefa de avaliar a segurança dos sistemas de IA a uma agência independente composta por especialistas, garantindo assim que as decisões sobre a segurança de sistemas de IA específicos sejam tomadas de forma isolada das pressões exercidas pela política eleitoral (Scherer, 2016, p. 393).

Essa tarefa essencial é atribuída às agências porque essas instituições estão, de acordo com Scherer, mais aptas do que os tribunais para avaliar a segurança de sistemas individuais de IA. Isso ocorreria, em grande parte, devido aos incentivos desalinhados do sistema judicial. Decisões relacionadas à segurança de uma tecnologia emergente não devem ser baseadas principalmente em depoimentos de especialistas contratados pelas partes em litígio, especialmente porque casos judiciais individuais raramente refletem os riscos gerais e benefícios associados a qualquer tecnologia (Scherer, 2016, p. 393).

Por fim, a AIDA aproveitaria a experiência dos tribunais na resolução de disputas individuais, atribuindo-lhes as tarefas de determinar se um sistema de IA se enquadra no escopo de um design certificado pela agência e de atribuir responsabilidade quando a interação entre múltiplos componentes de um sistema de IA resulta em danos (Scherer, 2016, p. 393).

Esse sistema sólido baseado em responsabilidade civil compeliria os projetistas e fabricantes a internalizar os custos associados aos danos causados pela IA, o que seria capaz de garantir a compensação para as vítimas e obrigando os projetistas, programadores e fabricantes de IA a avaliar a segurança de seus sistemas (Scherer, 2016, p. 393).

Para Scherer, o ponto de partida para regular a IA deve ser uma lei que estabeleça princípios gerais. A legislação criaria uma agência responsável por certificar programas de IA e classificá-los como seguros e estabeleceria os limites do poder da Agência para intervir na pesquisa e desenvolvimento de IA (Scherer, 2016, p. 395).

A AIDA deveria começar, como a maioria das leis modernas, com uma declaração de propósito. O propósito da AIDA seria garantir que a IA seja segura, protegida, suscetível ao controle humano e alinhada aos interesses humanos, tanto por meio da dissuasão da criação de IA que não possua essas características quanto por meio do estímulo ao desenvolvimento de IA benéfica, que inclua essas características (Scherer, 2016, p. 395).

A Agência seria obrigada a estabelecer regras definindo a inteligência artificial e as atualizar periodicamente. Regras relacionadas à definição de IA teriam que ser aprovadas pelo legislativo, pois essas regras efetivamente definem o escopo da jurisdição da Agência (Scherer, 2016, p. 395).

A AIDA daria à Agência a autoridade para estabelecer um sistema de certificação no qual sistemas de IA oferecidos para venda comercial poderiam ser avaliados por funcionários da Agência e certificados como sendo seguros. No entanto, em vez de proibir a IA não certificada, a AIDA funcionaria por meio de um sistema de responsabilidade civil dividido para incentivar projetistas e fabricantes a passarem pelo processo de certificação e, mesmo que optem por não obter a certificação, garantir a segurança e proteção de sua IA (Scherer, 2016, p. 395).

Sistemas que concluíssem com sucesso o processo de certificação da agência teriam uma responsabilidade civil limitada. Quando alguma ação judicial contra uma IA já certificada pela Agência fosse ajuizada, os autores dessas ações teriam que comprovar negligência real no design, fabricação ou operação de um sistema de IA para obter sucesso em uma ação de responsabilidade civil (Scherer, 2016, p. 395).

Se todas as entidades privadas envolvidas no desenvolvimento ou operação de um sistema de IA certificado pela Agência estivessem insolventes, um autor de ação judicial bem-sucedido teria a opção de apresentar uma reclamação administrativa à Agência para cobrir a falha, já que a Agência seria obrigada a administrar um fundo (financiado por taxas da Agência ou por apropriações do Congresso) suficiente para cumprir suas obrigações em caso de reclamações (Scherer, 2016, p. 395).

Sempre que uma ação judicial por negligência envolvendo o design de um sistema de IA certificado tivesse sucesso, a Agência seria obrigada a publicar um relatório semelhante aos relatórios que a Junta Nacional de Segurança no Transporte dos Estados Unidos prepara após acidentes e incidentes de aviação (Scherer, 2016, p. 395).

Já empresas que desenvolvessem, vendessem ou operassem IA sem obter a certificação da Agência seriam estritamente responsáveis pelos danos causados por essa IA. Além disso, a responsabilidade seria solidária, permitindo assim que um autor de ação judicial recupere o valor total de seus danos de qualquer entidade na cadeia de desenvolvimento, distribuição, venda ou operação da IA não certificada.

Um réu considerado responsável em tal ação judicial teria então que entrar com uma ação de regresso para obter restituição dos outros possíveis réus (Scherer, 2016, p. 395).

A Agência também seria obrigada a estabelecer regras para pesquisa e testes pré-certificação de IA. Essas regras permitiriam que desenvolvedores de IA coletassem dados e testassem seus projetos em ambientes seguros, para que a Agência pudesse tomar decisões de certificação com base em informações mais precisas (Scherer, 2016, p. 395).

Tais testes estariam isentos da responsabilidade civil rigorosa que normalmente estaria associada à IA não certificada. Além disso, a lei conteria uma cláusula que isentaria presumidamente programas em operação comercial doze meses antes da promulgação da lei, para evitar um impacto excessivo nas expectativas da indústria e dos consumidores (Scherer, 2016, p. 395).

No entanto, a AIDA deveria conceder à Agência a autoridade para criar um mecanismo separado do processo de certificação para revisar IA existente que possa representar um risco para o público (Scherer, 2016, p. 395).

Devido ao fato de a IA ser um assunto altamente técnico, os legisladores não estão adequadamente preparados para determinar quais tipos de IA representam um risco público. Portanto, eles devem delegar a tarefa de formular políticas substanciais de IA à uma agência composta por especialistas em IA com experiência acadêmica e/ou industrial relevante. Além das regras estabelecidas nos parágrafos anteriores, a AIDA daria à Agência a autoridade para especificar ou esclarecer a maioria dos aspectos do quadro regulatório de IA, o que inclui o processo de certificação da Agência (Scherer, 2016, p. 395).

De acordo com Scherer, a nova agência contaria com duas divisões: formulação de políticas e certificação. O órgão responsável pelas políticas teria o poder de definir IA (embora a definição precisasse ser aprovada pelo legislativo), criar exceções permitindo pesquisas em IA em certos ambientes sem a aplicação de rígida responsabilidade, e estabelecer um processo de certificação de IA (Scherer, 2016, p. 396-397).

O processo de certificação exigiria que desenvolvedores de IA que buscassem certificação realizassem testes de segurança e enviassem os resultados junto com o pedido de certificação para a agência. Os tomadores de decisão nas divisões de formulação de políticas e certificação deveriam ser especialistas com

formação ou experiência prévia em IA. O processo de contratação deveria garantir que a equipe de certificação incluísse uma combinação adequada de especialistas de acordo com as tendências predominantes em pesquisa em IA (Scherer, 2016, p. 396-397).

Na parte de formulação de políticas, a autoridade para criar regras ficaria a cargo de um Conselho. Como entidade administrativa independente, os membros do Conselho seriam indicados pelo poder executivo, sujeitos à aprovação do poder legislativo. Além de criar regras, o Conselho seria responsável por realizar audiências públicas sobre as regras e emendas propostas (Scherer, 2016, p. 396-397).

Na opinião de Scherer (2016, p. 396-397), provavelmente, a decisão política mais importante que a Agência enfrentaria seria como definir a inteligência artificial. Infelizmente, como mencionado anteriormente, a IA é um termo extremamente difícil de definir. Essas dificuldades tornam uma agência mais adequada para determinar uma definição funcional de IA para fins de regulamentação, uma vez que os legislativos e tribunais seriam certamente não teria competência técnica para estabelecer tal definição.

Sendo assim, seja qual for a definição adotada pela agência, ela deverá ser obrigada a revisar periodicamente e modificar essa definição, conforme necessário para refletir as evoluções na indústria (Scherer, 2016, p. 396-397).

A AIDA também exigiria que a Agência estabelecesse regras para testes prévios à certificação. Informações desses testes seriam um componente necessário em qualquer pedido de certificação para a Agência, e testes conduzidos em conformidade com as regras da Agência não estariam sujeitos a rígida responsabilidade.

As regras para esses testes seriam projetadas para garantir que eles sejam realizados em um ambiente controlado. Por exemplo, as regras poderiam proibir testes em computadores em rede, em robôs ou em outros sistemas com mecanismos que permitam a manipulação de objetos no mundo físico, em sistemas com poder computacional acima de um certo limite, ou em sistemas com quaisquer outras características que possam permitir que os testes de IA tenham efeitos fora do ambiente de teste (Scherer, 2016, p. 396-397).

A Agência teria a autoridade para acelerar as mudanças nos requisitos de teste. Essas mudanças entrariam em vigor imediatamente, mas também seriam

seguidas por um período de comentários públicos e uma votação subsequente para ratificar as mudanças (Scherer, 2016, p. 396-397).

Após a conclusão dos testes, os desenvolvedores de IA poderiam enviar um pedido de certificação à Agência. Para fornecer orientação aos solicitantes de certificação e estabelecer expectativas dentro da indústria, o Conselho seria responsável por determinar os critérios pelos quais as solicitações de certificação de IA seriam avaliadas (por exemplo, risco de causar danos físicos, alinhamento com objetivos e mecanismos para garantir o controle humano) (Scherer, 2016, p. 396-397).

A responsabilidade principal da equipe da Agência seria determinar se sistemas de IA específicos atendem a esses padrões. Empresas que buscam a certificação de um sistema de IA teriam que divulgar todas as informações técnicas relacionadas ao produto, incluindo: (1) o código-fonte completo; (2) uma descrição de todos os ambientes de hardware/software nos quais a IA foi testada; (3) como a IA se comportou nos ambientes de teste; e (4) quaisquer outras informações relevantes para a segurança da IA. Após a divulgação, a Agência realizaria seus próprios testes internos para avaliar a segurança do programa de IA (Scherer, 2016, p. 396-397).

Devido à diversidade de formas que a IA pode assumir, a Agência também teria o poder de limitar o alcance de uma certificação. Por exemplo, um sistema de IA poderia ser certificado como seguro apenas para uso em certos ambientes ou em combinação com determinados procedimentos de segurança (Scherer, 2016, p. 396-397).

A agência poderia criar um processo de certificação acelerado para sistemas de IA ou componentes que já foram certificados como seguros para uso em um contexto (por exemplo, veículos autônomos rodoviários) e que uma entidade deseja certificar como seguros para uso em um contexto diferente (por exemplo, aviões autônomos). Um processo de certificação igualmente simplificado seria criado para revisar e aprovar novas versões de sistemas de IA certificados (Scherer, 2016, p. 396-397).

A Agência também deveria criar regras que regulamentem as licenças e os requisitos de aviso para IA certificada. As regras poderiam especificar, por exemplo, que um projetista ou fabricante perderia a proteção de responsabilidade se vendesse um produto a um distribuidor ou varejista sem um acordo de licenciamento

que proíba esses vendedores de modificar o sistema de IA. Essa regra ajudaria a garantir que o produto que chega ao usuário final seja o mesmo produto certificado pela Agência (Scherer, 2016, p. 396-397).

Além da responsabilidade do poder legislativo e da agência a ser criada, Scherer também descreve o que, no ponto de vista dele, deveria ser o papel do poder judiciário. Sendo assim, para Scherer (2016), a responsabilidade dos tribunais dentro do arcabouço da AIDA seria resolver individualmente reclamações por danos causados pela IA, de modo a aproveitar a força institucional dos tribunais e sua experiência em apurar fatos.

De acordo com a estrutura de responsabilidade da AIDA, os tribunais, quando provocados, aplicariam as regras que regem ações por negligência em casos envolvendo IA certificada e as regras de responsabilidade estrita em casos envolvendo IA não certificada. Nessa última categoria de casos, a tarefa mais importante seria a alocação de responsabilidade entre os projetistas, fabricantes, distribuidores e operadores da IA causadora de danos. Em casos com vários réus e ações de indenização ou contribuição, a alocação de responsabilidade seria determinada da mesma forma que em casos comuns de delitos (Scherer, 2016, p. 398).

É bastante provável que, apesar do processo de certificação e dos requisitos de licenciamento, as partes em muitos casos questionem se a versão do sistema de IA em questão foi certificada pela Agência, ou disputem em que momento as modificações levaram a IA para fora do escopo das versões certificadas (Scherer, 2016, p. 398).

Em tais casos, o tribunal conduziria uma audiência antes do julgamento para determinar se o produto estava em conformidade com uma versão certificada do sistema no momento em que causou o dano e, caso não estivesse, identificaria o ponto em que o produto se desviou das versões certificadas. Esse ponto de modificação seria, então usado como a linha divisória entre os réus que têm responsabilidade limitada e aqueles que estão sujeitos à responsabilidade estrita (Scherer, 2016, p. 398).

Desta maneira, a inteligência artificial é uma tecnologia em rápido crescimento que tem o potencial de impactar profundamente diversos setores da sociedade, desde a economia e indústria até a saúde, educação e direito. Com o avanço acelerado da IA, é cada vez mais evidente a necessidade de uma regulação

adequada para lidar com os desafios e riscos associados a essa tecnologia inovadora. Em suma, os principais pontos que justificariam a regulação da inteligência artificial seriam:

1. Ética e Responsabilidade: A IA levanta questões éticas e morais importantes, especialmente quando se trata do uso de dados pessoais, tomada de decisões autônomas e potencial de substituição de empregos humanos. Uma regulação adequada pode estabelecer diretrizes e princípios éticos que orientem o desenvolvimento e o uso responsável da IA, garantindo que ela seja empregada para o benefício da sociedade como um todo.
2. Transparência e Responsabilização: A regulação da IA pode exigir que os sistemas de IA sejam transparentes e compreensíveis, tornando possível rastrear como as decisões são tomadas e identificar possíveis vieses ou preconceitos incorporados nos algoritmos. Além disso, a regulamentação pode estabelecer responsabilidades claras para os desenvolvedores e operadores de sistemas de IA em caso de danos ou violações de direitos.
3. Segurança e Privacidade: A IA lida com grandes volumes de dados, o que pode representar riscos significativos de violação de privacidade e segurança. Uma regulação adequada pode exigir padrões rigorosos de segurança cibernética e proteção de dados, garantindo que as informações dos usuários sejam tratadas de forma segura e confidencial.
4. Viés e Discriminação: Os sistemas de IA podem ser influenciados por vieses presentes nos dados de treinamento, o que pode levar a decisões discriminatórias ou injustas. A regulamentação pode estabelecer diretrizes para mitigar vieses e garantir que a IA seja justa e imparcial em suas decisões.
5. Impacto no Emprego: A automação impulsionada pela IA pode levar à substituição de empregos humanos em diversas indústrias. Uma regulação adequada pode ajudar a preparar a força de trabalho para essa transição, incentivando programas de requalificação e reconversão profissional.

6. Padrões de Segurança e Qualidade: A IA é usada em diversas aplicações críticas, como medicina, transporte e defesa. A regulamentação pode estabelecer padrões de segurança e qualidade para garantir que os sistemas de IA sejam confiáveis e precisos em suas tarefas.
7. Proteção do Consumidor: A regulamentação pode garantir que os consumidores estejam protegidos contra práticas comerciais desleais e enganosas relacionadas à IA, garantindo que eles tenham informações claras sobre como a tecnologia é usada e quais são seus benefícios e riscos.
8. Cooperação Internacional: Uma regulamentação consistente sobre a IA pode facilitar a cooperação internacional e promover o desenvolvimento de padrões globais para a tecnologia, evitando uma fragmentação regulatória que possa dificultar a inovação e o intercâmbio de conhecimentos.

A importância da regulação da inteligência artificial (IA) torna-se ainda mais crucial ao considerar a possibilidade de que a IA se desenvolva a ponto de ultrapassar a capacidade de controle humano. Essa perspectiva, conhecida como "superinteligência" ou "singularidade tecnológica", levanta preocupações profundas sobre os riscos potenciais e implicações éticas que podem surgir caso a IA alcance um nível de inteligência e autonomia superior à capacidade humana. Abaixo estão alguns pontos-chave que destacam a importância da regulação nesse cenário:

1. Riscos Existenciais: Se a IA se tornar superinteligente e ultrapassar a capacidade humana de compreendê-la e controlá-la, poderia representar um risco existencial para a humanidade. Uma IA descontrolada poderia tomar decisões que ameaçariam a própria existência da humanidade ou causariam danos irreversíveis ao planeta.
2. Governança e Tomada de Decisões: Uma IA altamente autônoma e superinteligente pode tomar decisões que afetariam de maneira significativa a sociedade, economia, política e questões globais. Nesse cenário, a regulação é essencial para garantir que a tomada de decisões da IA esteja alinhada com princípios éticos e valores

humanos, assegurando que o bem-estar da humanidade seja preservado.

3. Controle de Ações Autônomas: A regulação é necessária para definir limites e restrições sobre as ações autônomas da IA. Isso evitaria situações em que a IA possa ser usada indevidamente para fins prejudiciais ou antiéticos, protegendo a segurança e os interesses da sociedade.
4. Ética e Responsabilidade: Uma IA superinteligente pode ser capaz de desenvolver suas próprias regras e objetivos, o que levanta questões éticas e morais sobre sua autonomia e responsabilidade. A regulação deve estabelecer diretrizes claras para garantir que a IA seja projetada e utilizada de maneira ética, responsável e alinhada com os valores humanos.
5. Garantia de Benefícios Humanos: A regulação da IA é crucial para garantir que seu desenvolvimento e aplicação estejam voltados para beneficiar a humanidade como um todo. Regulamentações bem elaboradas podem orientar a pesquisa e o desenvolvimento da IA em direção a soluções que atendam às necessidades humanas e sociais, além de assegurar que a tecnologia seja usada para melhorar a qualidade de vida.
6. Proteção contra Abusos: A regulação pode proteger contra abusos e usos inadequados da IA. Com a possibilidade de IA superinteligente ser usada para fins maliciosos ou prejudiciais, uma estrutura regulatória sólida pode impedir seu uso indevido e garantir que a tecnologia seja aplicada de forma segura e benéfica.
7. Cooperação Global: Dada a natureza global da IA e os desafios associados a sua regulação, a cooperação internacional é essencial para estabelecer padrões e normas comuns. A colaboração entre países e organizações pode ajudar a mitigar os riscos da IA superinteligente e garantir uma abordagem unificada para questões críticas.

A regulação sobre a inteligência artificial é essencial para garantir que essa tecnologia seja desenvolvida e usada de forma ética, responsável e segura. Ao

estabelecer diretrizes e padrões claros, a regulação pode ajudar a maximizar os benefícios da IA para a sociedade, ao mesmo tempo em que mitiga os riscos e desafios associados a essa poderosa ferramenta. Uma abordagem equilibrada e colaborativa para a regulamentação da IA é fundamental para garantir que essa tecnologia contribua para o progresso humano de maneira sustentável e inclusiva.

1.2 Diferenciação entre Inteligência Artificial Discriminativa e Generativa

A pesquisa em Inteligência Artificial tem se deparado com abordagens distintas que refletem diferentes filosofias e estratégias. Duas dessas abordagens fundamentais são a IA Discriminativa e a IA Generativa. Este capítulo se dedica a elucidar as nuances e características que as diferenciam, proporcionando uma compreensão mais profunda sobre como essas modalidades abordam a tarefa de modelar dados e realizar tarefas específicas.

A IA Discriminativa direciona seus esforços à tarefa de mapeamento de entradas para categorias ou rótulos. Sua função consiste em aprender a fronteira de decisão que delimita distintas classes de dados. Esta abordagem é notável por sua eficiência, particularmente em contextos nos quais a discriminação direta entre classes é o objeto de interesse. Exemplos importantes de algoritmos discriminativos incluem Máquinas de Vetores de Suporte (SVM), Regressão Logística e Redes Neurais de camadas densas.

Ao contrário da perspectiva discriminativa, a IA Generativa concentra-se na modelagem da distribuição de probabilidade conjunta dos dados de entrada e saída. Esta abordagem busca compreender e capturar a estrutura subjacente dos dados, em contraste com a mera delineação de fronteiras de decisão. Modelos generativos destacam-se em contextos nos quais a criação de dados sintéticos ou a compreensão da relação intrínseca entre variáveis é imperativa. Exemplares de algoritmos generativos compreendem Redes Neurais Adversariais (GANs), Modelos Ocultos de Markov (HMMs) e Redes Neurais Recorrentes (RNNs).

A diferenciação entre IA Discriminativa e Generativa está na natureza de suas tarefas. Enquanto a primeira concentra-se em operações diretas de discriminação e classificação, a segunda almeja compreender a estrutura latente dos dados, possibilitando a geração de instâncias inéditas. Ambas as abordagens são

complementares e, juntas, oferecem um espectro abrangente de técnicas para abordar desafios diversos na IA.

Um importante exemplo de IA Generativa é o ChatGPT, que se assemelha a um "conversador" que, fundamentado em uma vasta base de dados, é capaz de gerar respostas novas e contextuais, estabelecendo-se como um agente capaz de criar no domínio da linguagem.

1.2.1 Chat GPT

O Chat GPT (*Generative Pre-trained Transformer*) é uma tecnologia de processamento de linguagem natural (PLN) desenvolvida pela OpenAI. Ele é baseado no modelo de linguagem GPT-3.5, que é uma das versões mais avançadas da família GPT.

A história da ferramenta remonta ao ano de 2018, quando o modelo original, o GPT-1, foi lançado pela OpenAI, que é uma organização de pesquisa em inteligência artificial que objetiva desenvolver tecnologias avançadas e promover o uso da IA e, por isso, ao longo dos anos vem desenvolvendo formas de aprimorar o conhecimento humano sobre essa tecnologia.

O GPT-1 foi um marco significativo na pesquisa em PLN, pois foi um dos primeiros modelos a utilizar uma arquitetura de rede neural, denominada Transformer, que se mostrou altamente eficiente para tarefas de processamento de linguagem natural. No entanto, mesmo com seus 117 milhões de parâmetros, o GPT-1 tinha limitações e não alcançava um desempenho excepcional.

Em resposta a essas limitações, a OpenAI continuou a aprimorar o modelo e lançou o GPT-2 em 2019. O GPT-2 era muito mais poderoso, com 1,5 bilhão de parâmetros e uma capacidade impressionante de gerar texto coeso e semelhante ao escrito por humanos. Devido ao seu tamanho e ao potencial risco de uso mal-intencionado, a OpenAI decidiu inicialmente não liberar completamente o GPT-2, mas disponibilizou versões menores para a comunidade de pesquisa.

Em 2020, a OpenAI deu mais um passo e lançou o GPT-3. Esse modelo foi um avanço em termos de escala, com 175 bilhões de parâmetros, tornando-se o maior modelo de linguagem de seu tempo. O GPT-3 surpreendeu muitos especialistas em IA com sua capacidade de realizar tarefas complexas de PLN e

produzir texto altamente persuasivo e quase indistinguível do produzido por humanos em muitos casos.

Com base no sucesso do GPT-3, a OpenAI continuou a melhorar o modelo, resultando no GPT-3.5. Essa versão representa uma interação mais recente do GPT-3 e possui um desempenho ainda mais aprimorado, eliminando algumas das limitações do modelo anterior.

O objetivo por trás da criação do Chat GPT, com base no GPT-3.5, é permitir que os usuários interajam com uma poderosa ferramenta de IA que possa responder a perguntas, fornecer informações e realizar tarefas de linguagem natural de forma mais eficiente. No entanto, a implementação do Chat GPT também vem com a conscientização de desafios éticos e de segurança, buscando equilibrar a utilidade da tecnologia com a proteção da privacidade e a mitigação de riscos potenciais associados ao seu uso.

Mas quais seriam as fontes utilizadas pelo Chat GPT? Esse modelo é pré-treinado através de uma enorme quantidade de textos de várias fontes da internet, que fazem com que haja um aprendizado e um entendimento das regularidades e dos padrões da linguagem. Isso permite que o GPT-3.5 compreenda contextos complexos e gere respostas coerentes e relevantes para as perguntas e solicitações dos usuários.

Os dados utilizados para treinar o GPT-3.5 além de buscar fontes da internet, também utiliza livros, artigos e fóruns online para montar sua base. Esses dados são usados para ensinar ao modelo as nuances e variações da linguagem humana e capacitá-lo a responder de maneira mais eficaz e natural às consultas dos usuários.

Outro ponto relevante possui relação com a proteção dos dados compartilhados pelos usuários. De acordo com a OpenAI, são tomadas medidas rigorosas para garantir a privacidade e segurança das informações e o Chat GPT é projetado para não armazenar dados pessoais de forma persistente após a interação com um usuário. Isso significa que as interações do usuário com o modelo são geralmente tratadas como transientes e não são mantidas em um banco de dados a longo prazo.

A partir de setembro de 2021, a OpenAI armazena os dados de interação com o modelo do GPT-3 por um período de 30 dias. Isso significa que, após 30 dias, as informações fornecidas pelos usuários durante a interação com o modelo deveriam ser excluídas e não seriam mantidas nos servidores da OpenAI.

Segundo a OpenAI, essa medida foi adotada como parte dos esforços da OpenAI para garantir a privacidade e a segurança dos dados dos usuários e para minimizar o risco associado ao armazenamento de informações pessoais ou sensíveis. Ao limitar o tempo de retenção dos dados, a OpenAI buscaria mitigar possíveis riscos de vazamentos de informações confidenciais e proteger a privacidade dos usuários. Além disso, a OpenAI garante que possui políticas claras para o tratamento de dados sensíveis e informações confidenciais.

No entanto, apesar dos esforços para proteger a privacidade dos usuários, de acordo com a própria OpenAI, o uso do Chat GPT também apresenta alguns riscos e desafios, que seriam:

1. Vazamento de informações confidenciais: Embora a OpenAI tenha implementado medidas de segurança, pode haver riscos de vazamento de informações confidenciais durante uma interação, especialmente se o usuário compartilhar dados pessoais ou sigilosos com o modelo.
2. Viés de linguagem: Como o GPT-3.5 é treinado com dados da internet, ele pode refletir vieses e estereótipos presentes nesses dados. Isso pode levar a respostas que não são neutras e perpetuam desigualdades sociais.
3. Respostas incorretas ou enganosas: Embora o GPT-3.5 seja poderoso, ele não é infalível e pode gerar respostas incorretas ou enganosas, especialmente quando apresentado com informações inexatas ou ambíguas.
4. Uso inadequado: O Chat GPT pode ser explorado por usuários mal-intencionados para disseminar desinformação, realizar phishing ou conduzir atividades ilegais.

Atualmente, além dos riscos relacionados a privacidade de dados, confidencialidade e segurança das informações, existe ainda uma preocupação proveniente do uso do Chat GPT para fornecimento de informações muito técnicas, como por exemplo, consultas médicas ou jurídicas. Nesse momento, o risco de não se ter uma regulação sobre o assunto deixa de ser somente um receio de que a inteligência artificial possa fugir do controle humano e passa a preocupar pelo medo de que as pessoas adotem a inteligência artificial como principal fonte de

conhecimento, o que pode trazer um conjunto de informações não tão assertivas para a sociedade.

A interpretação das normas requer uma análise profunda do texto, do contexto histórico e das decisões judiciais pertinentes. A inteligência artificial pode enfrentar dificuldades na compreensão desses aspectos mais complexos da interpretação jurídica, limitando-se apenas à análise superficial de palavras e padrões em documentos. A falta de capacidade de compreender detalhes e argumentos jurídicos pode levar a respostas imprecisas e incompletas, prejudicando a qualidade das consultas técnicas.

A interpretação jurídica é uma atividade complexa que envolve a análise cuidadosa de textos legais, princípios constitucionais, precedentes judiciais e outros elementos contextuais. Essa tarefa requer uma compreensão profunda da linguagem jurídica e das normas constitucionais. Embora a inteligência artificial tenha avançado nos últimos anos, ainda existem limitações significativas em relação à interpretação jurídica, que precisam ser consideradas:

1. Contexto e História: A interpretação jurídica muitas vezes depende do contexto histórico e do propósito por trás de uma norma constitucional. A IA pode ter dificuldade em compreender a importância do contexto e da evolução histórica de certas disposições. Além disso, ela não tem capacidade de extrair significado de eventos históricos ou sociais que influenciaram a redação das leis.
2. Intenção do Legislador: A interpretação jurídica frequentemente envolve considerar a intenção do legislador ao criar uma determinada lei ou norma constitucional. A IA não possui consciência ou compreensão da intenção humana, o que a limita na análise de debates legislativos, discussões parlamentares e outros registros que possam indicar a intenção do legislador.
3. Analogia e Argumentos jurídicos: Advogados e juristas frequentemente usam analogias e argumentos jurídicos sofisticados para fundamentar suas interpretações. Essas técnicas exigem a compreensão do contexto, da cultura jurídica e das tradições. Embora a IA possa reconhecer padrões, ela não tem a capacidade de compreender o

significado por trás desses padrões e, portanto, pode não ser capaz de fazer analogias ou aplicar argumentos complexos.

4. Princípios Éticos e Filosóficos: A interpretação jurídica muitas vezes envolve a aplicação de princípios éticos e filosóficos que não são facilmente quantificáveis ou codificáveis em algoritmos. Questões como justiça, equidade e bem comum são complexas e subjetivas, tornando-se desafiador para a IA fornecer respostas completas e satisfatórias.
5. Mudanças na Sociedade: O direito constitucional evolui para refletir as mudanças na sociedade e os valores emergentes. A IA, no entanto, depende de dados históricos e pode ter dificuldade em se adaptar a novas realidades sociais, resultando em respostas que podem estar desatualizadas ou inadequadas para o contexto atual.
6. Uso de Precedentes: A utilização de precedentes judiciais é uma prática comum na interpretação jurídica. A IA pode analisar grandes volumes de casos passados, mas pode ter dificuldades em compreender as diferenças sutis entre casos e identificar quais precedentes são mais relevantes para uma determinada questão constitucional.

Embora a inteligência artificial ofereça benefícios em agilizar a pesquisa e análise de informações jurídicas, suas limitações na interpretação jurídica não podem ser negligenciadas. As complexidades e aspectos subjetivos envolvidos na atividade de interpretação ainda exigem o julgamento humano, experiência e compreensão das particularidades do Direito. A IA pode ser uma ferramenta útil para apoiar e complementar as pesquisas jurídicas, mas é fundamental que os profissionais do Direito estejam cientes de suas limitações e mantenham o papel central do raciocínio humano na tomada de decisões jurídicas fundamentais para a sociedade.

Tratando também das decisões jurídicas, quando a IA é utilizada para auxiliar em pesquisas jurídicas, suas conclusões podem influenciar a tomada de decisões de profissionais do Direito e até mesmo de tribunais. No entanto, a responsabilidade pela decisão final ainda reside nos humanos que utilizam a tecnologia. Erros ou imprecisões nas análises feitas pela IA podem levar a consequências sérias, especialmente em questões constitucionais sensíveis. Portanto, é essencial garantir

que os profissionais compreendam as limitações da IA e a utilizem como uma ferramenta de apoio, não como substituta do julgamento humano.

A tomada de decisões jurídicas é uma das principais funções do sistema de justiça e envolve a análise cuidadosa de fatos, leis, precedentes e princípios constitucionais para resolver conflitos e questões legais. Essa atividade requer um alto grau de discernimento e interpretação por parte dos profissionais do direito, que precisam considerar diversos aspectos ao tomar uma decisão. Embora a inteligência artificial tenha demonstrado avanços significativos em algumas áreas, existem várias limitações quando se trata de tomar decisões jurídicas, tais como:

1. Discricionariedade e Equidade: Em muitos casos, a lei pode ser ambígua ou permitir espaço para interpretação e discricionariedade. A IA é baseada em algoritmos e regras predefinidas, o que a torna menos adequada para lidar com questões que exigem equidade, senso de justiça e considerações éticas. A tomada de decisões jurídicas frequentemente envolve o equilíbrio entre valores concorrentes, e a IA pode não ser capaz de pesar essas considerações de forma adequada.
2. Contexto e Conhecimento Prévio: As decisões jurídicas muitas vezes são influenciadas por conhecimentos específicos do caso, contexto histórico e outros fatores que podem não estar disponíveis para a IA. A compreensão do cenário mais amplo e das nuances do caso é fundamental para tomar uma decisão bem fundamentada, e a IA pode não ser capaz de incorporar esses elementos em sua análise.
3. Tomada de Decisão em Tempo Real: Em algumas áreas do direito, como por exemplo nos casos criminais, decisões precisam ser tomadas rapidamente em situações de alta pressão. A IA pode levar tempo para analisar e processar informações, o que pode ser inadequado em contextos de tomada de decisões em tempo real.
4. Relevância dos Precedentes: A IA pode analisar grandes volumes de casos passados e precedentes, mas pode ter dificuldades em entender a relevância e o contexto específico de cada precedente para um caso em particular. A jurisprudência é frequentemente baseada em analogias e considerações específicas, que requerem um julgamento humano mais aprofundado.

5. Falta de Empatia e Sensibilidade Humana: A tomada de decisões jurídicas pode envolver empatia e compreensão das circunstâncias pessoais das partes envolvidas. A IA não tem a capacidade de compreender plenamente a complexidade das emoções humanas, o que pode ser importante em alguns casos para alcançar decisões justas.

Embora a IA tenha mostrado potencial em muitas áreas, incluindo a análise de dados e a pesquisa, ainda existem desafios significativos em relação à sua aplicação na tomada de decisões jurídicas. A complexidade do sistema legal, a importância do contexto e do conhecimento prévio, bem como a necessidade de discernimento ético e equidade, tornam a tomada de decisões jurídicas uma atividade intrinsecamente humana. A IA pode servir como uma ferramenta de apoio para os profissionais do direito, mas é fundamental que eles continuem a desempenhar um papel central no processo decisório, garantindo que os valores constitucionais e a justiça sejam preservados na sociedade.

1.3 Projeto de lei nº 2338 de 2023

A necessidade de regulação dos sistemas que utilizam Inteligência Artificial é algo uníssono dentre os estudiosos do assunto, principalmente pelos riscos que essas tecnologias podem trazer para a raça humana. Por esse motivo, alguns países vêm impulsionando os debates e, cada vez mais, caminham para a criação de legislações objetivando a criação de regras mais claras e de responsabilização pelos atos praticados pela IA.

No caso do Brasil, em maio de 2023 foi publicado um Projeto de Lei de iniciativa do Senador Rodrigo Pacheco que tem como foco estabelecer normas gerais para o desenvolvimento, implementação e uso responsável de sistemas de inteligência artificial no Brasil. Esse Projeto se baseia na proteção dos direitos fundamentais e na garantia à implementação de sistemas seguros e confiáveis, em benefício da pessoa humana, do regime democrático e do desenvolvimento científico e tecnológico (Brasil, 2023).

Primeiramente, imperioso destacar que a iniciativa de regulação é extremamente importante e inovadora, demonstrando que o Brasil está atento às

novas evoluções tecnológicas e preparando-se para que, de maneira mais segura, a IA seja cada vez mais aplicada no país. Através da leitura do Projeto de Lei, é possível perceber que ele se baseia em alguns fundamentos, tais como: respeito aos direitos humanos, não discriminação, desenvolvimento tecnológico e inovação, livre concorrência, proteção de dados, acesso à informação e promoção da pesquisa e do desenvolvimento.

Além desses fundamentos, o Projeto baseia-se ainda em alguns princípios, como: (1) desenvolvimento sustentável; (2) liberdade de decisão e escolha; (3) transparência; (4) devido processo legal; e principalmente, (5) participação humana no ciclo da inteligência artificial; (6) supervisão humana efetiva; (7) auditabilidade; (8) confiabilidade; (9) responsabilização; e (10) reparação integral dos danos.

Aparentemente, em primeira análise, os princípios listados pelo Projeto de Lei parecem completamente adequados e justificados, entretanto, quando se inicia um estudo mais aprofundado do tema, resta claro que existe uma lacuna, que ainda necessita de preenchimento.

No âmbito do “dever-ser”, a participação humana, supervisionando a atuação da inteligência artificial é primordial e viável, assim como a responsabiilização e reparação integral dos danos. Todavia, quando se observa a prática e o funcionamento dessas ferramentas, percebe-se que em diversos momentos, a velocidade e a capacidade de obtenção de resultados de um sistema como esses será muito superior a capacidade de um ser humano.

Desta maneira, esperar que um ser humano sinta-se confortável e seja capaz de contestar e supervisionar um software que analisa em segundos uma quantidade enorme de dados, interpretando-os e criando solução de maneira ágil e, em grande parte das vezes, com boa acuracidade, é algo complexo. Ao mesmo tempo, a responsabilização e reparação dos danos, apesar de importantíssima, também deve levar em consideração a grande quantidade de fases de construção de um sistema desse tipo, contando com uma vasta categoria de profissionais e com atualizações constantes de software, dificultando a escolha de quem deveria ser o indivíduo responsável, como se demonstrará no Capítulo 3.

Sendo assim, a listagem de princípios contida no artigo 3º do referido projeto de lei, apesar de pertinente, parece superficial, carecendo, portanto, de maiores elucidações de conceitos e explicações de como, na prática, atingir todos esses objetivos será viável (Brasil, 2023). Apesar de ser sabido que as leis não

necessariamente trazem a solução para os problemas enfrentados pela sociedade, tendo também como um dos principais papéis demonstrar quais seriam os comportamentos esperados e considerados ideais, publicando-se uma lei com uma redação que tenha vácuos e que demonstre desconhecimento da realidade dos sistemas culminaria com a criação de um texto legal obsoleto e sem aplicabilidade.

O Projeto de Lei traz ainda algumas definições bastante pertinentes, tanto no âmbito prático, quanto conceitual, expondo o que seriam, do ponto de vista desta lei, sistemas de inteligência artificial, fornecedor de sistema de inteligência artificial, operador de sistema de inteligência artificial, agentes de inteligência artificial, autoridade competente, discriminação e mineração de textos e dados. Ao contrário do que Scherer recomendaria, o projeto brasileiro não inclui a definição de inteligência artificial, porém, essa definição pode ser extraída mediante interpretação da definição de sistema de inteligência artificial, sendo possível concluir que, nacionalmente, estaria-se definindo-a como algo baseado em aprendizado de máquina, que é capaz de demonstrar conhecimento por meio de dados e tendo como objetivo produzir previsões, recomendações ou decisões.

Considerando toda a discussão existente sobre a definição de inteligência artificial e o fato de não necessariamente ela partir de um aprendizado de máquina ou ter como objetivo previsão ou recomendação, seria válido que, de maneira mais clara, houvesse a inclusão de um significado mais técnico e embasado em conclusões de estudiosos do assunto.

Um dos pontos mais interessantes também abordados é a avaliação preliminar, que consiste basicamente em uma classificação de risco do sistema, a ser realizada pelo próprio fornecedor e ratificada, ou não, pela autoridade competente. Todavia, ainda ficam pendentes muitas explicações sobre como deve ser realizada essa classificação, quais serão os critérios, qual o momento em que deve ser submetido, o que será avaliado pela autoridade.

Vale ressaltar que a inteligência artificial pode ser usada como complemento a alguns sistemas, ou seja, o sistema não tem como objetivo final realizar uma previsão, mas em alguma das etapas de seu funcionamento, alguma das ferramentas utilizadas utiliza essa tecnologia. Nesse momento, como ficaria a avaliação preliminar? Nos dias atuais, a quantidade de softwares que em algum momento utilizam IA é enorme e, caso uma autoridade competente tenha que

autorizar o funcionamento de cada um, o princípio do desenvolvimento e da pesquisa pode ficar prejudicado.

O artigo 17 do Projeto de Lei (Brasil, 2023) descreve o que seria considerado como sistema de inteligência artificial com alto risco, demonstrando uma grande preocupação com decisões que podem conter vieses e, de alguma forma, serem discriminatórias. Todavia, não há qualquer menção a sistemas que não necessitam dos humanos para continuar atuando, ou seja, aqueles mais autônomos e, consequentemente com maior potencial danoso para a sociedade a longo prazo. O que existe, na verdade, é uma avaliação de impacto algorítimo, que só é realizada naqueles sistemas considerados de alto risco.

Como é sabido, a responsabilidade aplicada aos sistemas é um dos pontos mais importantes quando o assunto é inteligência artificial. Sendo assim, as expectativas sobre o capítulo que trata desse assunto no Projeto de Lei são altas, visto que a responsabilidade pode impactar diretamente nos riscos do negócio e nos riscos dos profissionais.

Apesar da expectativa, no que tange o Capítulo V – Responsabilidade Civil (Brasil, 2023), a redação não atende às necessidades do mercado e da sociedade e, mais uma vez, demonstra ser rasa e sem muito conhecimento técnico do tema. A opção do legislador nesse caso foi por responsabilizar os fornecedores ou operadores, caso algum dano seja sofrido e divide essa responsabilização em dois grupos:

1. Sistema de alto risco: fornecedor ou operador respondem somente pelos danos causados, na medida da sua participação no dano.
2. Sistema que não seja de alto risco: a culpa do agente causador do dano é presumida.

Nesse momento, alguns pontos merecem destaque. O primeiro é o fato dos sistemas de alto risco não possuirem responsabilidade objetiva do fornecedor. Isso pode se dar em razão da existência de uma avaliação prévia mais detalhada realizada pela agência reguladora. Todavia, mesmo a agência tendo realizado a avaliação, não há qualquer menção a responsabilização da mesma em caso de erro que cause dano tanto ao fornecedor, quanto a terceiro. Além disso, no momento em que a responsabilização ocorre na medida da proporção do dano, o usuário, que

muitas vezes possui uma hipossuficiência técnica passa a ter que comprovar a culpa, o que à priori não parece adequado.

Vale ressaltar ainda que estão previstas algumas possibilidades onde fornecedores e operadores não serão responsabilizados, que seriam quando: “I- comprovarem que não colocaram em circulação, empregaram ou tiraram proveito do sistema de inteligência artificial; ou “comprovarem que o dano é decorrente de fato exclusivo da vítima ou de terceiro, assim como de caso fortuito externo” (Brasil, 2023).

Desta forma, algumas dúvidas surgem, como por exemplo: como seria possível um sistema de inteligência artificial ser utilizado e causar danos sem que o fornecedor ou operador o tenha colocado em circulação? Como o fornecedor comprovaria que não colocou um sistema em circulação? O que seria colocar um sistema em circulação? O que seria considerado como fortuito externo em casos de sistema de inteligência artificial? Os tribunais brasileiros estão preparados para lidar com argumentos técnicos em casos sobre esse assunto?

É possível compreender também a partir da análise, que as pessoas físicas seriam consideradas como parte integrante e, portanto, sob responsabilidade do fornecedor. Entretanto, deve-se levar em consideração que as empresas podem buscar direito de regresso e, nesse momento será necessário compreender quem foi o culpado pelo equívoco, tarefa extremamente complexa e relevante, apesar do silêncio do Projeto sobre o assunto.

2 INTELIGÊNCIA ARTIFICIAL E A PRÁTICA JURÍDICA

A sociedade mundial passa constantemente por atualizações, através da disseminação do aprimoramento de alguns processos, que acaba por alterar a cultura e provocar revoluções no mercado de trabalho e nos hábitos dos indivíduos. A primeira revolução industrial, por exemplo, foi capaz de fazer com que fosse utilizado o vapor e água como mecanismo, ao invés da mão de obra humana e animal; já a segunda, introduziu a produção em massa, com linhas de montagem, fazendo com que uma quantidade menor de trabalhadores fosse necessária e que suas atividades fossem mais mecanizadas.

Após muitos anos, adentrou-se à era dos computadores e, mais uma vez, a forma com que os seres humanos viviam foi drasticamente modificada, configurando o que se pode denominar terceira revolução industrial. Entretanto, parece que hoje, se está testemunhando a quarta revolução industrial, protagonizada pela internet das coisas e inteligência artificial, uma vez que, é possível perceber a conectividade entre os mais diversos aparelhos tecnológicos e a alta evolução dos aplicativos capazes de tomar decisões de maneira similar a de um indivíduo.

Apesar dos grandes benefícios que essa quarta revolução pode trazer, os danos sociais não devem ser olvidados. Atualmente, estão disponíveis no mercado jurídico diversos escritórios de advocacia, que disputam suas posições e sobrevivência diariamente, seja mostrando competitividade através do preço, seja através da qualidade do serviço. Todavia, a partir do momento que a inteligência artificial é introduzida, o quesito tecnologia passa a ser mais um quesito agregado à competitividade. Abaixo, trecho da obra Inteligência Artificial e Direito tratando das previsões e da forma com que os advogados devem adaptar-se, confira-se:

> [...] previsões sobre o desenvolvimento futuro da inteligência artificial são tão confiantes quanto diversas". Portanto, é o momento de se ter cautela, mas observando os desdobramentos. De qualquer modo, existem projeções sobre o avanço da chamada "inteligência de máquina de nível humano", que representa "aquela que pode executar a maioria das profissões humanas ao menos tão bem quanto um humano típico": 10% de probabilidade de que esta modalidade de inteligência artificial seja alcançada até 2022, 50% que isto seja atingido até 2040, e 90% de probabilidade até 2075. Os dados apontam tendências de evolução da inteligência artificial. Como não se tem uma pesquisa quantitativa abrangente, estes indicativos deverão ser analisados com cuidado, e se deveriam fazer pesquisas aprofundadas. Entretanto, pela pesquisa bibliográfica realizada se verifica que existem grandes chances de estas tendências se tornarem realidade. Mesmo que se

> concretizem parcialmente, já causarão impactos profundos no trabalho das pessoas. Cabe ao Direito estudar quais serão os impactos nas diversas carreiras jurídicas, buscando antecipar cenários e projetar estruturas normativas flexíveis capazes de acompanhar a evolução tecnológica, com o devido suporte regulatório (Engelmann; Werner, 2020, p. 141).

Desta forma, depreende-se que o direito como um todo, não se excluindo, nesse caso, os magistrados, promotores, procuradores, professores e doutrinadores, que, dentro de seus nichos de atuação devem buscar atualizar-se e, desde já, preparar-se para as inovações que virão. Far-se-á importante destacar, portanto, que a gradação das mudanças é primordial para que o impacto social seja menor, cabendo a comunidade jurídica colaborar.

Temos a nossa disposição atualmente *Big Data*, Inteligência Artificial e *Robotic Process Automation* (RPA), que com certeza poderiam contribuir para a construção de uma reformulação.

> Através da utilização de algoritmos, é possível que consigamos respostas consideradas subjetivas e que claramente envolvem juízo de valor, podendo, de algum modo, substituir o pensamento humano, que muitas vezes é influenciado por fatores externos. A matéria-prima utilizada pelos algoritmos para tais decisões é o Big Data, ou seja, a enorme quantidade de dados disponíveis no mundo virtual que, com o devido processamento, pode ser transformada em informações economicamente úteis, que servirão como diretrizes e critérios para o processo decisório algorítmico.

Não obstante, antes de adentrarmos especificamente no tratamento e exploração das espécies tecnológicas, cabe primeiramente defini-las.

Entende-se como Inteligência Artificial a capacidade de obter, através de análises de dados, respostas subjetivas e mais próximas a que um humano forneceria. Nesse caso, o chamado Machine Learning seria o responsável por, por meio de algoritmos, compreender o padrão lógico normalmente utilizado no universo de dados acessado. Desta forma, através da Inteligência Artificial e do Machine Learning, temos a oportunidade de obter de maneira ágil, a tendência de retorno para determinada necessidade.

> A Inteligência Artificial usa a informação externa obtida por esses meios como um input para a identificação de regras e modelos subjacentes ao confiar em perspectivas, como o aprendizado das máquinas, o qual descreve métodos que auxiliam os computadores a aprenderem sem serem explicitamente programados. Entretanto, a Inteligência Artificial é mais ampla que o próprio aprendizado das máquinas, uma vez que também cobre a habilidade de um sistema de perceber os dados ou de controlar, mover e manipular objetos, com base nas informações aprendidas, seja por

> meio de um robô, seja por outro dispositivo conectado. Os sistemas de Inteligência Artificial podem ser atualmente classificados como analíticos, inspirados em humanos e humanizados. Os primeiros dispõem de características da inteligência cognitiva, criam representações do mundo e usam o conhecimento com base em experiências passadas para a tomada de decisões futuras. Já os segundos possuem elementos de inteligência cognitiva e emocional, inclusive emoções na tomada de decisão, enquanto os últimos reúnem competências de inteligência cognitiva, emocional e social e são capazes de ter consciência própria nas interações com outros (Steibel; Vicente; Jesus, 2019, p. 41).

Os juristas Gustavo Tepedino e Rodrigo da Guia no artigo que escreveram sobre inteligência artificial e suas consequências, demonstram certa resistência ao uso desta tecnologia exatamente no que diz respeito ao Machine Learning. Para eles, o fato dos resultados serem proveniente do aprendizado da máquina em cima de uma atividade humana anterior, pode vir a causar insegurança e a gerar resultados equivocadas, confira-se:

> Nesse contínuo - e potencialmente ilimitado - processo de autoaprendizagem, tende a se incrementar gradativamente a complexidade das interações desenvolvidas por tais sistemas autônomos. Quanto mais livres (i.e., não supervisionadas ou controladas) as experiências, maior o grau de imprevisibilidade dos aprendizados e dos atos a serem praticados. Como já se identificou em doutrina, verifica-se relação inversamente proporcional entre a influência do criador e a influência do ambiente no desenvolvimento do sistema. Intensifica-se, assim, a aptidão dos sistemas sob exame à tomada autônoma de decisões e à produção de resultados que não poderiam ser efetivamente previstos pelos seus programadores – e tampouco pelos usuários diretos (Tepedino; Silva, 2019, p. 73).

Vale ressaltar que existe certa quantidade de formas utilizadas para que as máquinas adquiram conhecimento. A obra Inteligência Artificial e Direito explica muito bem o tema, confira-se:

> Na contemporaneidade, um elemento definidor dos sistemas de Inteligência Artificial é a multiplicidade de formas, a partir das quais se dá a habilidade de aprendizado pelas máquinas com base em informações passadas. Podem-se identificar três tipos de processos de aprendizado primordiais: o supervisionado, o não supervisionado e o aprendizado reforçado. O primeiro mapeia um conjunto de inputs para um dado conjunto de resultados, incluindo métodos como regressão linear, árvores de classificação e redes neurais. No segundo, os inputs são rotulados, mas os resultados não, o que significa que o algoritmo precisa inferir a estrutura subjacente dos próprios dados, como na análise de clusters, que visa a agrupar elementos em categorias similares, mas nas quais nem a estrutura dos clusters nem seu número são conhecidos antecipadamente. Os usuários precisam colocar uma confiança maior no sistema de Inteligência Artificial. Isso ocorre, por exemplo, no reconhecimento do comando de voz dado pelo usuário. No aprendizado reforçado, o sistema recebe um resultado variável para ser

> maximizado e uma série de decisões que podem ser tomadas para impactá-lo (Kaplan; Haenlein, 2018, p. 2).

Sendo assim, a partir do explicado acima, é possível compreender que a Inteligência Artificial adquire papel importante em alguns processos, quais sejam, organização de dados, auxílio na tomada de decisão e automação da decisão (Steibel; Vicente; Jesus, 2019, p. 43). Os três processos podem de certa forma se completar, de modo que a organização dos dados permite que o profissional seja capaz de analisar de maneira assertiva as informações a respeito de seu negócio, auxiliando na tomada de decisão, escolhendo, assim, os melhores caminhos a serem seguidos. Nos últimos tempos, os relatórios gerencias, materializados de modo que o entendimento seja facilitado, possibilitam que grandes decisões sejam tomadas de maneira mais segura e com base em números confiáveis.

Todavia, alguns questionamentos e inseguranças ainda pairam pela mente daqueles que estudam o tema e acompanham os resultados dos seus experimentos. Considerando a forma com que a inteligência artificial funciona, não seria possível que a mesma perpetuasse os preconceitos inerentes à raça humana? A ética, que muitas vezes falta aos seres humanos, seria imputada nos sistemas, fazendo com que decisões integras sejam tomadas, alterando drasticamente a humanidade do direito e da sociedade como um todo?

Essas perguntas se mostram presentes no quotidiano de todos, devido ao fato de ainda existir e ser noticiado pela mídia brasileira e internacional casos de uso indevido dos dados dos clientes, havendo quebra do dever ética, de princípios constitucionais, como privacidade e dignidade da pessoa humana. Desta forma, é primordial o trabalho conscientizador, no sentido de fazer com que inclusão, transparência, ética, segurança e *accountability* estejam enraizados, ou, pelo menos, se tornem parte de normas reguladoras capazes de impor punições severas para aqueles que as descumprirem. A obra Inteligência Artificial e Direito dá destaque especial à *accountability*, conforme se verifica:

> Este último princípio, o da accountability, talvez seja um dos mais importantes e que merecem uma atenção especial. Sem tradução exata para a língua portuguesa, trata-se de um conceito do idioma inglês que abarca práticas que remetem à reponsabilidade com ética, à obrigação, à busca por transparência, à prestação de contas. De maneira simplificada, significa que aqueles que desempenham funções relevantes na sociedade deveriam dar transparência ao que estão fazendo, por quais motivos, e como estão fazendo. Remete à necessidade de governança e, em alguns

> casos, até de responsabilidade civil. Não é à toa que o tema da IA tem chamado a atenção de empresas, governos e organizações nacionais e internacionais (Gutierrez, 2019, p. 61).

Andreas Kaplan, citado pela obra Inteligência Artificial e Direito, acredita que podem ser elencados três tipos de Inteligência artificial: restrita, geral e superinteligência (Steibel; Vicente; Jesus, 2019, p. 44).

O primeiro nível, ou seja, a inteligência restrita, utilizada por todos durante as pesquisas no Google, por exemplo, é aquela capaz de maneira ágil e incomparável, apresentar eficiência em tarefas específicas, substituindo e superando a capacidade humana, mas que ainda não atingiu maturidade para atividades mais complexas.

O segundo nível, que seria a inteligência geral, tornou-se capaz de compreender determinada informação, contextualizá-la corretamente e, a partir daí, tomar decisões. Sendo assim, a inteligência geral se mostra bastante robusta, tendo como foco e principal resultado almejado, o alcance da similaridade com a intuição humana, possuindo plenas condições de tomar decisões como se se tratasse da consciência de um ser humano.

A junção do primeiro nível com o segundo nível, cumulada com uma evolução ainda maior nos sistemas, poderiam originar em algo sem precedentes, totalmente revolucionário e capaz de alterar de maneira drástica a vida de todos. Algo como a superinteligência, com habilidade de interagir socialmente, inovar e de maneira autônoma criar novas soluções, poderia de certo modo romper paradigmas e criar grandes oportunidades para a sociedade.

A expectativa com essa nova tecnologia, vem em conjunto com o medo das consequências que a mesma pode trazer. Vale ressaltar, todavia, que o bom uso e o tratamento cauteloso devem sempre prevalecer, visto que o uso desorientado pode aumentar ainda mais a desigualdade social e o abismo econômico já existente.

O Big Data seria, portanto, a capacidade de coletar e armazenar informações e harmonizá-las, de modo a possibilitar uma análise das mesmas, gerando insights aptos a alterar a estratégia das empresas, determinando com maior exatidão riscos e causas de falhas, reduzindo custos e proporcionando tomadas de decisão mais embasadas.

É exatamente o efeito do *Big Data* a causa da implementação da Lei Geral de Proteção de Dados. Os dados devem, portanto, ser mapeados e estar disponíveis para seus donos, quando desejarem a ter acesso aos mesmos. A possibilidade de

exclusão, consulta e restrição pelos usuários das informações a serem disponibilizadas diminui a probabilidade de ilícitos e torna capaz a responsabilização dos infratores.

Já o RPA, diferentemente dos conceitos explicados acima, funciona através de robôs capazes de automatizar tarefas manuais e repetitivas, que normalmente exigem investimento de homem-hora, mas que na verdade são tarefas operacionais, onde o lado intelectual é pouco exigido. O software, portanto, consegue reproduzir por conta própria e em altíssima velocidade, exatamente as mesmas atividades que um ser humano executaria, sem, no entanto, produzir qualquer juízo de valor.

Sendo assim, a grande distinção entre RPA e inteligência artificial seria a profundidade da automação gerada. O risco acoplado a ambas também é bastante distinto, visto que com a parametrização adequada dos dados, a probabilidade de um equívoco ocorrer no trabalho realizado por robôs torna-se categoricamente inferior.

Nessa conjuntura, temos sistemas capazes de, dentre outras funcionalidades, compreender decisões judiciais, criando padrões de resposta de determinado magistrado, relacionando-o aos mais diversos tipos de pedido. Desta forma, para cada necessidade do cliente, temos a probabilidade de êxito e as principais formas de argumentação, em um procedimento denominado jurimetria.

Vale ressaltar ainda que, o cálculo do contingenciamento é um dos principais motivos de embate entre advogados e departamentos financeiros das Companhias. Existem dois fatores que geram esse atrito, quais sejam: subjetividade do Direito, que é capaz de gerar resoluções distintas para pedidos semelhantes, fato que ocorre pela subjetividade do Direito, que não é ciência exata e, de modo algum, deve ser totalmente positivista; e falha na padronização da classificação dos riscos.

É inegável que o Direito é uma ciência humana e que é exatamente nesse ponto que se encontra sua beleza e exuberância. A capacidade de mudança de paradigmas e de divergência de opiniões é a mais nobre demonstração da democracia e isso deve, sim, permanecer independente do transcorrer dos anos. Contudo, através da observação do histórico de julgados de um mesmo magistrado, é possível obter uma estimativa das chances de êxito, mesmo que ainda com possibilidade de variação. Essa estimativa mais exata e gerada por uma máquina, com certeza traz maior segurança aos departamentos financeiros, torna as

informações mais transparentes para os investidores e eleva ainda mais a imagem dos departamentos jurídicos.

Entretanto, apesar de todos os benefícios que a Inteligência Artificial pode trazer aos juristas, há quem discorde da grandiosidade da revolução proporcionada. Isso porque nós, como seres humanos que somos, não tomamos atitudes baseadas totalmente nas fontes do direito, mas também em nossas subjetividades e experiências passadas, o que não poderia ser alcançado pelos algoritmos. Sendo assim, Harry Surden, professor associado da Universidade de Colorado, coloca na introdução de um de seus brilhantes artigos que, há quem realmente acredite que até que as máquinas sejam capazes de reproduzir impressões de um nível tão alto quanto a de advogados treinados e experientes, o impacto causado será ínfimo.

Gustavo Tepedino e Rodrigo da Guia (2019) parecem discordar dessa possibilidade e demonstram total discordância e receio com as consequências que as novas tecnologias trariam para o sistema jurídico brasileiro. Em artigo tratando do tema, ambos parecem não acreditar no potencial de evolução e assertividade dos resultados trazidos pela inteligência artificial, o que somente prejudica ainda mais o processo de atualização das ciências jurídicas:

> Essa tão advertida imprevisibilidade repercute também na definição do que exatamente deve ser considerado falha no funcionamento do código de programação (ou, simplesmente, *bug*, na sintética formulação do inglês já consagrada na práxis). A figura-se tênue, com efeito, a linha divisória entre o dano (que se espera não previsto, em homenagem à presunção de boa-fé subjetiva) produzido por sistema autônomo defeituoso e o dano produzido por sistema autônomo não defeituoso. Em meio às dúvidas sobre o que se deveria considerar sistema defeituoso, cresce não apenas o potencial de lesão à coletividade exposta às novas tecnologias, mas também o temor da responsabilização de uma pessoa por danos imprevisíveis causados pelos sistemas autônomos (Tepedino; Silva, 2019, p. 73).

A quarta revolução industrial precisa chegar ao Direito brasileiro, que claramente, necessita da crença na inovação e da atualização daqueles que representam a mais alta cúpula do direito nacional. Postergar esse momento, não testar essas novas formas de vivenciar a prática jurídica e desincentivá-las, é a mais clara demonstração de desconhecimento, de ausência de consciência social, de extremo individualismo e de protecionismo à classe de advogados.

Observando-se as atuais formas de utilização de *Machine Learning*, como por exemplo a tradução automática de textos, é incontestável que, apesar de existirem imperfeições no trabalho produzido, o ganho é enorme e muitas das vezes, os erros são aceitáveis. Sendo assim, mesmo que se tenha pequeno equívocos, o resultado obtido com a automação de atividades jurídicas, certamente, é incontestável.

Vale ressaltar ainda que há quem defenda a existência de duas formas distintas de *Machine Learning*, a supervisionada e a não supervisionada:

> Por sua vez, os sistemas baseados em Machine Learning possuem maior grau de complexidade. Apesar de também serem programados a priori, sua construção algorítmica é feita de maneira a aprenderem com a interação com um ambiente externo dinâmico e a partir dela fazerem correlações e reconhecerem padrões. Uma diferença marcante entre o Machine Learning para os algoritmos de análise simples é que o primeiro é capaz de analisar, fazer correlações e buscar padrões a partir de dados não estruturados: fotos, vídeos, textos, dados coletados por smartphones e sensores. De uma maneira simples, podemos classificar os sistemas de Machine Learning em dois subgrupos. Aqueles que são supervisionados dos não supervisionados. O aprendizado de máquina supervisionado é aquele no qual os critérios de correlações iniciais são parametrizados (ou “ensinados”) por seres humanos. Em ambientes dinâmicos, são necessárias várias interações iniciais de “treinamento”/“calibragem” do sistema de IA por um humano com domínio naquele contexto específico até que o sistema consiga resultados mais precisos e minimamente satisfatórios. Por sua vez, o Machine Learning não supervisionado consegue dispensar essa calibragem inicial por seres humanos. Isso é alcançado por meio do desenvolvimento de novas tecnologias, como as redes neurais ou o deep learning, que são capazes de criar padrões de correlações próprios, alheios ao raciocínio humano. Isso é alcançado por meio da criação de uma rede de múltiplas unidades não lineares de processamento de dados que se retroalimentam de modo a imitar (de maneira rudimentar) um cérebro humano. Esses sistemas de IA são capazes de analisar um ambiente dinâmico e dele extrair correlações e padrões por si só (Gutierrez, 2019, p. 61).

É imperioso frisar ainda que, o Direito, como sabemos, é baseado em diversos fatores, que vão além da legislação vigente, ou seja, as fontes do direito são compostas por elementos como contexto social em que está inserida a sociedade, moral do legislador e suas vivências anteriores, jurisprudências, dentre outros fatores. Nesse sentido, considerando que a inteligência artificial tem como premissa a compreensão das atitudes humanas e a sua reedição, partindo de um histórico de atividades semelhantes realizadas, paira sobre a nova tecnologia uma insegurança com relação ao futuro das decisões e das argumentações.

A sociedade altera de maneira constante a sua forma de visualizar as situações diárias, os costumes se atualizam e, consequentemente, legislações se tornam obsoletas e carecem de modernização. As sentenças proferidas pelos

tribunais brasileiros devem refletir o que na atualidade é compreendido como justo, permitindo que o direito caminhe em conjunto com os seus usuários, de modo a preservar a justiça.

Desta forma, fica um questionamento: a inteligência artificial seria capaz de não somente perpetuar argumentos, decisões e modelos contratuais baseadas no histórico, mas também de promover as atualizações necessárias aos novos riscos e costumes de uma comunidade que evolui a todo tempo?

A resposta para essa pergunta ainda parece um pouco cinzenta, se observado o momento da inteligência artificial e do *machine learning*, uma vez que, apesar das grandes vitórias construídas, ainda há um *gap* no que diz respeito as previsões e a percepção de riscos externos

Para que um aplicativo seja capaz de estar atemporal, será necessário que uma outra vertente de modo de operação seja posta em prática, sendo capaz de acessar fontes diversas, externas ao mundo jurídico, percebendo tendências humanas e aplicando-as de maneira independente. Entretanto, enquanto ainda não se tem todo esse alcance e desenvoltura tecnológica, é plausível que os humanos, como parte imprescindível no universo jurídico interfiram nessa transação e parametrizem as máquinas de acordo com o mundo em que está inserido.

A jurimetria, se tratada em conjunto com o RPA, que é capaz de, por meio de parametrizações realizar as mais diversas tarefas manuais como por exemplo, redigir peças processuais de contencioso cível, bem como gerar contratos padrão, pode otimizar o tempo dos advogados e tornar o serviço oferecido a seus clientes mais transparente, exato e de qualidade reconhecida.

As ferramentas hoje disponíveis, são capazes de gerar indicadores gerenciais sobre contratos, desempenho dos advogados e escritórios de advocacia, além de funcionar como repositório de documentação, de disponibilizar o valor a ser contingenciado e realizar pesquisas reputacionais de possíveis parceiros, clientes e fornecedores. Sendo assim, o benefício vai muito além de automação e consegue atingir o compliance das organizações.

Um ponto crucial seria a avaliação das chances de êxito para os casos. Normalmente, a realizamos com base na experiência na advocacia, legislação, pesquisa jurisprudencial e doutrinária, bem como estilo do tribunal e magistrado responsável pelo julgamento. Estudos como esse, para que tenham excelência e

atinjam as expectativas do cliente, demandam tempo elevado dos profissionais e, ainda assim, geram discussões intermináveis entre os envolvidos.

O uso de *Machine Learning* e *Big Data* nesses casos, pode sim, auxiliar-nos, tanto por trazer com exatidão e rapidez o que exigiria de um humano muita dedicação, quanto por ter como plano de fundo uma ferramenta confiável e tecnológica.

Contudo, apesar de todos os benefícios citados acima, ainda existem diversos obstáculos a serem superados, como por exemplo o da alimentação dos sistemas em formato padrão, parametrização e adaptação à nova conjuntura, bem como o custo. Obstáculos esses que fazem parte do processo natural e inevitável de evolução e construção de um status quo superior para o futuro.

2.1 O uso de tecnologia aplicado ao Compliance e aos contratos

O direito, em sua essência, constitui-se como meio para resolução de conflitos tendo como base em regramentos pré-estabelecidos. Entretanto, com o advento da tecnologia, a realidade indubitavelmente será alterada:

> Esse cenário está passando por rápidas transformações: o surgimento de sistemas, de algoritmos, viabilizadores da inteligência artificial estão sendo desenvolvidos pelo próprio ser humano para tomar decisões, avançando das mais simples às mais complexas. E mais. Sob certas condições, esta inteligência artificial aprende e tem condições de aprender sozinha. Ao mesmo tempo, o Direito, como área de conhecimento, sempre esteve assentado nos pressupostos da certeza, segurança e previsibilidade. Tudo isto está em transformação, caracterizando o que Ulrich Beck chama de "metamorfose do mundo", ou seja, "ao invés de mudança, metamorfose, que desestabiliza as certezas da sociedade moderna, os eventos e processos que provocam um choque fundamental; [...] a metamorfose significa que o que foi impensável ontem é real e possível hoje". (Engelmann; Werner, 2020, p. 141).

Matéria publicada na Revista Exame estima que cerca de 30% das vagas atuais, serão ocupadas por robôs. Desta forma, à medida que a tecnologia evolui, a qualidade dos serviços prestados melhora e a oferta de trabalho diminui. Nesse sentido, torna-se inevitável a atualização dos profissionais, para que estejam atuando em sintonia com o mercado.

O Compliance vem sendo uma área bastante observada e difundida pelos juristas e pela sociedade como um todo. Os casos de corrupção ao redor do mundo

e os prejuízos, tanto financeiros, quanto de imagem fizeram com que as companhias passassem a investir mais tempo e dinheiro na instituição de uma cultura de ética e integridade nos negócios.

No Brasil, corrupção é, com certeza, o assunto mais tratado nos noticiários, sendo também, motivo dos mais intensos debates e de grandes discórdias e bipolaridade. Isso porque os cidadãos toleram cada vez menos atos antiéticos e passam a ter cada vez mais consciência acerca das atitudes diárias do contexto social em que está envolvido. A mudança de cultura é, portanto, um passo necessário para que a sociedade mundial se torne mais integra, menos individualista e mais humana.

Todavia, diferentemente do que muitos possam imaginar, Compliance não existe somente para regular e fiscalizar fraudes capazes de tornarem-se escândalos de grandes proporções. O Compliance é, primordialmente, o responsável pela alteração no modus operandi na Companhia, ou seja, aquele que deve atentar-se para as relações interpessoais, a diversidade, a igualdade de gênero e a integridade do colaborador como ser humano antes do que como profissional.

Sendo assim, a forma com que os dados dos clientes são tratados por uma sociedade empresária depende muito grau de integridade dos seus colaboradores, que precisam compreender e estar engajados em uma cultura de combate ao uso inadequado das informações recebidas, num ambiente de pleno respeito a proteção da dignidade da pessoa humana e da privacidade.

Os diferentes monitoramentos, auditorias e a gestão dos riscos de Compliance, por esse motivo, possuem um papel importantíssimo num contexto social de revolução tecnológica, onde a responsabilidade social sob as consequências que a inteligência artificial e o *machine learning* podem trazer tornam-se parte da realidade diária e estão, a todo tempo, sendo observadas pelos titulares dos dados e pelos clientes, que também se encontram mais criteriosos com relação a esse ponto.

Ademais, alguns dos processos internos de auditoria, monitoramento e gestão de riscos também podem e devem utilizar-se da inteligência artificial e do RPA para realizarem suas verificações, sem, no entanto, dispensar a análise de um profissional preparado para analisar seus resultados. Quando isso ocorre, os trabalhos de verificação de toda a documentação referente a procedimentos, bem como a análise dos pagamentos realizados pela Companhia, por exemplo, podem

ser concluídos com agilidade, reportando instantaneamente transações suspeitas e que podem estar relacionados a corrupção, fraude, suborno, cartel, uso indevido de dados de clientes para obtenção de melhores resultados, dentre outros tipos de benefícios.

Quando a tecnologia, através da inteligência artificial, torna-se capaz de compreender o comportamento de um colaborador com a intenção de cometer alguma violação aos preceitos éticos ensinados por meio do *machine learning*, a quantidade de casos de corrupção tende a cair ou a tornar-se cada vez mais rara, visto que os níveis de dificuldade crescem à medida que os controles passam a ser mais precisos.

Desta forma, o Compliance é claramente importante para que haja o uso consciente da tecnologia. Os departamentos jurídicos, como grandes impactados positivamente pela inteligência artificial podem aprimorar sua atuação de diversas formas, como poderá ser verificado adiante.

Muito se fala da utilização da tecnologia para o Direito Processual, todavia, o Direito Contratual também pode ser bastante beneficiado. Os instrumentos contratuais, em sua maioria, são baseados em princípios corporativos, que originam diferentes modelos de documentos a serem utilizados a depender do tipo de prestação de serviço.

Para essas situações, podemos utilizar tanto o RPA, quanto a Inteligência Artificial. O primeiro, pode ser aplicável a situações como contratos de adesão, onde são alteradas somente os dados das partes. Já a IA, utilizada em contratos mais complexos, pode ser capaz de auxiliar na interpretação das cláusulas, percebendo a ausência de cláusulas consideradas pétreas e entendendo quando há incongruência com os princípios corporativos.

Trata-se, portanto, da figura dos contratos “inteligentes”, ou seja, aqueles capazes de automatizar a averiguação, a aplicação das principais cláusulas e sua execução. Edna Hogemann trata brilhantemente do assunto, em citação à Vermeulem, conforme se verifica:

> Claramente, os contratos inteligentes se tornarão mais predominantes no crescente mundo da Internet das Coisas. Quanto mais dispositivos estiverem conectados uns aos outros, mais “contratos inteligentes” serão usados para processar e fazer cumprir as “transações legais”. A tecnologia Blockchain pode ajudar a tornar as transações verificáveis e seguras. Um blockchain é um banco de dados ou banco de dados digital compartilhado

> que mantém uma lista crescente de registros de transações recentes entre as partes participantes, envolvendo dispositivos e ativos digitais. O blockchain garante a verdade, integridade e autenticidade das informações necessárias para entrar em transações de “contrato inteligente” (Vermeulem, 2018, p. 01 *apud* Hogemann, 2018, p. 109).

Desta forma, a ideia de se ter diversas tecnologias interligadas e conectando informações das mais diversas origens parece criar uma maneira de se lidar com o direito contratual. Uma área que já é consolidada como bastante importante para as empresas, passa a ser para os advogados, uma área menos manual e mais dependente de conhecimentos estratégicos para a composição e criação de novas cláusulas.

As chamadas “Legal Techs” nesse contexto assumem papel primordial, visto que são as criadoras dos softwares munidos de tecnologia suficiente para realizar a parametrização e utilizar os algoritmos necessários.

A aplicação dessas tecnologias acaba por garantir que sempre serão respeitadas as normas e a governança da empresa. Com a grande valorização do Compliance, possuir softwares capazes de por meio de parametrização condicionar a geração de um contrato a existência de clausulado anticorrupção e *Due Diligence*, por exemplo, é, certamente, um ponto a ser apreciado.

Até mesmo o *Due Diligence*, trabalho realizado através da leitura e análise de documentação, pode ser beneficiado pelas tecnologias. Atualmente, utilizamos o *Big Data* para reunir informações disponíveis e conectados pela Internet das Coisas, ao mesmo tempo que a Inteligência Artificial é capaz de selecionar esses dados e compreender quais configuram notícias negativas. A partir daí, selecionados os principais documentos, a própria ferramenta gera relatório com parecer final sobre a empresa e a relação entre as partes.

Sendo assim, profissionais preparados para analisar dados, aproveitam a quantidade de notícias negativas e todos os elementos oferecidos e, com facilidade, criam Dashboards com informações gerenciais e mapeamento de onde estão os principais riscos de Compliance, corroborando para que atitude sejam capazes de mitigar riscos de suborno.

Portanto, podemos concluir que a Internet da Coisas é uma realidade, onde todos os dados gerados são capazes de se conectar e alterar a concorrência nos mais diversos mercados, incluindo o advocatício. Inicialmente, é provável que tenhamos a impressão de maior aumento na desigualdade provocado pelo alto valor

envolvido para adquirir o software e o HH necessário para a preparação, padronização dos dados e parametrização.

Porém, com a difusão desse tipo de serviço, maior será a quantidade de fornecedores, bem como a variação na profundidade do uso da tecnologia, que pode ser adotada de maneira gradual, mediante alteração da cultura dos envolvidos no processo. Desta forma, resta clara a importância da mudança da mentalidade dos juristas e da sociedade como um todo, para que aos poucos se torne comum o entendimento de que a tecnologia é capaz de aumentar a qualidade dos serviços, diminuir o tempo de duração dos processos, melhorar o aceso à justiça, proporcionar que os advogados se dediquem a assuntos mais complexos, trabalhando com as informações trazidas pelos robôs e deixando os processos transparentes e adequados aos princípios de integridade.

A utilização dos benefícios tecnológicos é, desta forma, uma grande oportunidade de alavancar a prestação de serviços jurídicos, tornando-os mais eficientes, alinhados com o momento globalmente vivenciado, equilibrando as relações e criando oportunidades de crescimento.

3 RESPONSABILIDADE E INTELIGÊNCIA ARTIFICIAL

Tecnologias avançadas e serviços inovadores, incluindo inteligência artificial especializada para tarefas específicas, trazem consigo um grande potencial. Essas inovações já trouxeram benefícios consideráveis, especialmente ao aumentar a eficiência, precisão, pontualidade e conveniência em uma ampla gama de serviços digitais (Yeung, 2018, p. 14).

No entanto, o surgimento dessas tecnologias também despertou crescente preocupação pública quanto aos possíveis efeitos prejudiciais: tanto para indivíduos quanto para grupos vulneráveis e para a sociedade em geral. Para que essas tecnologias possam ser um elemento positivo, impulsionando o florescimento tanto individual quanto social, é crucial que se compreenda melhor essas preocupações. Isso não apenas demanda uma compreensão mais aprofundada de como essas tecnologias afetam o desfrute dos direitos humanos e liberdades fundamentais, mas também exige uma reflexão cuidadosa sobre a questão da responsabilidade diante das possíveis consequências negativas (Yeung, 2018, p. 14).

Por esse motivo, o Comitê de Especialistas em Dimensões dos Direitos Humanos do processamento automatizado de dados e diferentes formas de inteligência artificial, produziu um estudo sobre a responsabilidade relacionada ao uso de inteligência artificial. Este estudo parte do pressuposto de que, em sistemas democráticos e constitucionais, os conceitos, instituições e práticas relacionados à responsabilidade desempenham um papel fundamental. Isso é essencial para assegurar que tanto indivíduos quanto organizações sejam devidamente responsabilizados pelos efeitos prejudiciais de suas ações, além de estabelecer e manter as bases para uma cooperação social e coordenação pacíficas e confiáveis (Yeung, 2018, p. 14).

Portanto, o objetivo do estudo foi analisar as implicações das tecnologias digitais avançadas (incluindo a IA) no conceito de responsabilidade, especialmente no que se refere à possibilidade de prejudicar o desfrute dos direitos humanos e liberdades fundamentais protegidos pela Convenção Europeia dos Direitos Humanos, e como a responsabilidade por esses riscos e consequências deve ser distribuída (Yeung, 2018, p. 14).

A abordagem metodológica adotada pelo Comitê foi interdisciplinar, aproveitando conceitos e pesquisas acadêmicas das áreas do direito, humanidades,

ciências sociais e, em menor grau, ciência da computação. A conclusão a que eles chegaram é que, caso se deseje levar os direitos humanos a sério em uma era digital globalmente conectada, não se pode permitir que o poder das tecnologias digitais avançadas e sistemas, assim como dos indivíduos e entidades que as desenvolvem e implementam, seja acumulado e exercido sem considerar as consequências (Yeung, 2018, p. 14).

As nações têm a responsabilidade primordial de proteger os direitos humanos. Portanto, é necessário garantir que aqueles que criam, desenvolvem e se beneficiam dessas tecnologias sejam responsabilizados pelos impactos negativos. Isso inclui a obrigação de estabelecer mecanismos institucionais eficazes e legítimos que evitem as violações dos direitos humanos que essas tecnologias possam representar e que também cuidem da integridade do ambiente sócio-técnico coletivo no qual os direitos humanos e o Estado de Direito estão enraizados (Yeung, 2018, p. 14).

Para compreender as implicações da inteligência artificial no conceito de responsabilidade sob a ótica dos direitos humanos, é essencial adquirir um entendimento básico de como essas tecnologias são desenvolvidas e operam (Yeung, 2018, p. 14).

3.1 Inteligência das máquinas e aprendizado automático

Muito do entusiasmo em relação à promessa e ao potencial da IA para gerar avanços e melhorias em diversos domínios sociais, como produtividade industrial, saúde, medicina, gestão ambiental e segurança alimentar, depende da capacidade e potencial do aprendizado automático. O aprendizado automático é a tecnologia que possibilita que os computadores executem tarefas específicas de forma inteligente, aprendendo a partir de exemplos, dados e experiência (Yeung, 2018, p. 15).

Embora as técnicas de aprendizado automático já existissem há algum tempo, houve avanços significativos nos últimos anos devido a desenvolvimentos tecnológicos, aumento na capacidade de processamento e um aumento substancial na disponibilidade de dados digitais. Esses progressos permitiram o desenvolvimento de máquinas que agora podem superar os seres humanos em tarefas específicas (como processamento de linguagem, análise, tradução e

reconhecimento de imagens), algo que até poucos anos atrás era um desafio para elas alcançarem resultados precisos (Yeung, 2018, p. 15).

Essas tecnologias são agora onipresentes na vida cotidiana das pessoas que vivem em sociedades altamente industrializadas e contemporâneas. Nesses ambientes, as pessoas agora interagem regularmente com sistemas de aprendizado automático que permitem que serviços digitais (como motores de busca, sistemas de recomendação de produtos e sistemas de navegação) forneçam respostas precisas e eficientes às perguntas dos usuários em tempo real, enquanto melhoram constantemente seu desempenho aprendendo com os erros cometidos (Yeung, 2018, p. 15).

3.2 Propriedades relevantes para a responsabilidade da IA

Para compreender como as tecnologias digitais avançadas desafiam as noções existentes de responsabilidade legal, moral e social é crucial identificar os atributos ou propriedades relevantes para a responsabilidade que essas tecnologias possuem, ou seja, as características dessas tecnologias que provavelmente afetarão como elas impactam os outros (Yeung, 2018, p. 15).

3.2.1 Automatização de tarefas

Com esse propósito, uma das características mais significativas dessas tecnologias reside na sua habilidade de realizar tarefas (muitas das quais anteriormente exigiam operadores humanos) de forma "automática", ou seja, sem precisar da intervenção direta de seres humanos (Yeung, 2018, p. 15).

3.2.2 Autonomia das máquinas

Os avanços nas técnicas de machine learning resultaram no desenvolvimento e uso cada vez maior de sistemas que não apenas são automatizados, mas que operam de maneira autônoma. Embora o termo "autonomia" seja frequentemente usado para descrever muitas aplicações habilitadas para IA em discussões públicas e políticas, dentro da comunidade técnica não parece haver um consenso amplamente aceito sobre o que exatamente esse termo significa e quais são as

condições necessárias para caracterizar uma entidade não humana como "autônoma" (Yeung, 2018, p. 15).

No entanto, de acordo com o estudo do Comitê, em algumas fontes literárias o termo "autonomia" frequentemente é usado para se referir à capacidade funcional de agentes computacionais de executar tarefas de forma independente, exigindo que o agente tome "decisões" sobre seu próprio comportamento sem entrada direta de operadores humanos e sem controle humano. Agentes computacionais desse tipo operam ao perceber o ambiente ao seu redor e adaptar seu comportamento em resposta ao feedback sobre o desempenho de suas próprias tarefas, de forma que suas decisões e ações não são consideradas "totalmente determinísticas" desde o início (e, portanto, não totalmente previsíveis com antecedência) devido à grande variedade de contextos e ambientes nos quais esses agentes podem operar. Assim compreendida, a autonomia é uma característica que pode estar mais ou menos presente em graus, dependendo do nível de supervisão e intervenção humana necessária para a operação do sistema (Yeung, 2018, p. 15).

Como explicado anteriormente, em capítulos anteriores, alguns sistemas que utilizam machine learning se destacam por sua habilidade de aprender e se transformar ao longo do tempo, ajustando dinamicamente seus próprios sub-objetivos, e sua capacidade de se adaptar às condições locais por meio de informações de sensores externos ou dados de entrada atualizados (Yeung, 2018, p. 19).

Os criadores individuais do sistema podem escolher e configurar seu estado inicial e parâmetros, incluindo o objetivo principal que pretendem otimizar, mas uma vez implantado, o funcionamento e os resultados do sistema evoluirão com o uso em diferentes ambientes (Yeung, 2018, p. 19).

De acordo com o Comitê de Especialistas em Dimensões dos Direitos Humanos do processamento automatizado de dados e diferentes formas de inteligência artificial, esses sistemas computacionais têm a intenção de operar de maneira que permita ao sistema tomar decisões independentes, que escolham entre alternativas que não são pré-programadas com antecedência, e fazem isso sem qualquer intervenção humana. Os sistemas de IA atuais não conseguem determinar o objetivo principal para o qual o sistema é projetado para otimizar (o que deve ser especificado pelos desenvolvedores humanos do sistema), mas eles têm a

capacidade de determinar seus próprios sub-objetivos intermediários (Yeung, 2018, p. 19).

Além de sua capacidade de operar sem supervisão e controle humano diretos, essas tecnologias possuem várias outras características relevantes para a responsabilidade, que seriam algumas das características destacadas anteriormente, com por exemplo inescrutabilidade e opacidade, complexidade, dinamismo, input humano, interação, discrição, escalabilidade, conectividade global, operação contínua, capacidade de gerar ideias através de base de dados, dentre outras (Yeung, 2018, p. 19).

A relevância de compreender as facetas dos direitos humanos no contexto da IA é evidente nas várias investigações e relatórios encomendados e produzidos por um número crescente de organizações da sociedade civil. Além disso, essa temática tem se tornado cada vez mais central nas pesquisas acadêmicas voltadas para a "ética da IA". Isso abrange o trabalho do Conselho da Europa, incluindo seu estudo sobre as implicações dos direitos humanos nas técnicas de processamento automatizado de dados e nas possíveis implicações regulatórias. Esse estudo, realizado pelo Comitê de Especialistas em intermediários da internet (MSI-NET), é conhecido como "Estudo Wagner" (Yeung, 2018, p. 19).

O Estudo Wagner identifica exemplos de sistemas de tomada de decisão algorítmica atualmente em uso que podem violar ou prejudicar o gozo de "direitos que estão claramente em discussão pública, variando em grau de impacto". Esses direitos incluem o direito a um julgamento justo e ao devido processo legal, o direito à privacidade e à proteção de dados, a liberdade de expressão, a liberdade de associação, o direito a um recurso efetivo, a proibição da discriminação e o direito a eleições livres (Yeung, 2018, p. 20).

Entretanto, como concluiu o Relatório encomendado pela Assembleia Parlamentar do Conselho da Europa e realizado pelo Rathenau Instituut:

> Apesar do amplo impacto das tecnologias digitais nos direitos humanos, até agora, houve pouca atenção dada a esse tópico crucial e quase nenhum debate político e público substancial a respeito. Como resultado, está ocorrendo uma séria erosão dos direitos humanos. Portanto, o debate sobre direitos humanos, que está notavelmente atrasado em relação aos desenvolvimentos tecnológicos acelerados, precisa ser fortalecido de maneira urgente (Yeung, 2018, p. 40).

O Comitê de Especialistas em Dimensões dos Direitos Humanos do processamento automatizado de dados e diferentes formas de inteligência artificial se baseia no Estudo Wagner ao analisar criticamente como as tecnologias digitais avançadas podem estar relacionadas ao conceito de responsabilidade. O Estudo Wagner adota uma "perspectiva de direitos humanos", focando em como essas tecnologias podem minar a capacidade prática de exercer determinados direitos humanos e liberdades de maneira sistemática em uma época dominada pelas tecnologias avançadas de IA. Isso é feito ao invés de se aprofundar na análise detalhada de aplicativos de IA específicos que possam afetar adversamente direitos humanos e liberdades fundamentais particulares (Yeung, 2018, p. 19).

São consideradas duas dimensões desses impactos sistemáticos: primeiro, as ameaças a um conjunto de direitos impostas por sistemas de tomada de decisão algorítmica. Em segundo lugar, os efeitos sociais coletivos adversos mais amplos das tecnologias de IA (incluindo, mas não se limitando àquelas incorporadas aos sistemas de tomada de decisão algorítmica), dos quais apenas alguns podem ser prontamente expressos na linguagem do discurso existente sobre direitos humanos. Com o tempo, esses efeitos adversos mais amplos poderiam ameaçar sistematicamente as bases sócio-técnicas que fundamentam a própria noção de direitos humanos e nas quais eles estão enraizados (Yeung, 2018, p. 22).

De acordo com o estudo do Comitê, muitos analistas afirmam que progressos em tecnologias digitais interconectadas, incluindo aquelas atualmente denominadas tecnologias de IA, estão impulsionando o surgimento de uma 'Nova Revolução Industrial' que provocará mudanças abrangentes em todos os aspectos da vida social, de uma magnitude e alcance que serão tão perturbadoras e desestabilizadoras quanto aquelas causadas pela Revolução Industrial original. Antes de examinar as ameaças e riscos potenciais associados a essas tecnologias emergentes, o Comitê entendeu ser útil destacar brevemente o contexto sócio-político e econômico que permeia essa nova realidade (Yeung, 2018, p. 22).

Observando esse contexto, o Comitê entendeu que pode haver paralelos entre os impactos sociais mais amplos da revolução industrial original e os impactos esperados da 'Nova' Revolução Industrial que está agora se iniciando. Por exemplo, embora a Revolução Industrial do século XIX tenha trazido inúmeros benefícios tanto para indivíduos quanto para a sociedade, e possa ser creditada com melhorias

substanciais e disseminadas nos padrões de vida e no bem-estar individual e coletivo, ela gerou efeitos adversos não intencionais (Yeung, 2018, p. 22).

Isso inclui tanto efeitos adversos diretos na saúde e segurança humana associados às formas iniciais de produção industrial quanto a queima de combustíveis fósseis para alimentar a atividade industrial, o que levou a um grave problema de mudança climática em escala global e que ainda não foi adequadamente abordado ou resolvido. No entanto, os efeitos adversos na mudança climática decorrentes das tecnologias que provocaram a Revolução Industrial original só se tornaram evidentes após mais de um século, quando já era tarde demais para lidar com eles de maneira eficaz (Yeung, 2018, p. 22-23).

As sociedades contemporâneas podem agora enfrentar um dilema semelhante. Uma das dificuldades em tentar identificar e prever os maiores efeitos adversos na sociedade da inovação tecnológica decorre não apenas das dificuldades em prever suas possíveis aplicações e adoção, mas especialmente das dificuldades em antecipar seus efeitos combinados e acumulativos ao longo do tempo e do espaço (Yeung, 2018, p. 22-23).

3.3 A ascensão dos sistemas de tomada de decisão algorítmica

Sistemas computacionais que utilizam algoritmos de aprendizado de máquina, combinados com a rápida e ampla adoção de dispositivos 'inteligentes', têm estimulado o surgimento de sistemas de tomada de decisão algorítmica que buscam aproveitar (e frequentemente monetizar) os dados que agora podem ser obtidos rastreando sistematicamente e coletando os vestígios digitais deixados pelos comportamentos on-line dos indivíduos, e usando tecnologias avançadas (incluindo IA) para gerar novos conhecimentos que podem ser usados para embasar decisões do mundo real (Yeung, 2018, p. 22-23).

Muitos desses sistemas dependem de técnicas orientadas por dados que envolvem a coleta sistemática e em massa de dados de uma população de indivíduos para identificar padrões e, assim, prever preferências, interesses e comportamentos de indivíduos e grupos, frequentemente com graus muito altos de precisão. Esses perfis de dados podem então ser usados para classificar indivíduos com o objetivo de identificar 'candidatos de interesse' visando gerar 'percepções acionáveis' - ou seja, percepções que podem ser usadas para embasar e

automatizar a tomada de decisões sobre indivíduos por parte daqueles que realizam o perfilamento (ou seus clientes) (Yeung, 2018, p. 22-23).

Esses sistemas são amplamente utilizados por varejistas que buscam direcionar produtos para indivíduos identificados como os mais lucrativos e mais propensos a se interessar por eles, por atores políticos e organizações que buscam adaptar e direcionar mensagens de campanha para indivíduos que são identificados como os mais propensos a serem persuadidos por elas, e, cada vez mais, por autoridades do sistema de justiça criminal que buscam avaliar o 'risco' que determinados indivíduos são identificados de maneira algorítmica como representando para a segurança pública, a fim de tomar decisões de custódia sobre esses indivíduos (sejam suspeitos criminais ou aqueles condenados por infrações criminais) (Yeung, 2018, p. 22-23).

É nesse contexto sócio-econômico que surgiram preocupações públicas sobre os efeitos sociais das tecnologias digitais avançadas (incluindo IA) (Yeung, 2018, p. 22-23).

3.3.1 Como esses sistemas ameaçam sistematicamente direitos específicos

O uso de sistemas de tomada de decisão algorítmica pode ameaçar sistematicamente vários direitos, como por exemplo: (a) o direito a um julgamento justo e o devido processo legal; (b) o direito à liberdade de expressão; (c) o direito à privacidade e à proteção de dados; e (d) a proibição da discriminação no gozo de direitos e liberdades.

3.3.1.1 *O direito a um julgamento justo e o devido processo legal*

Muitos sistemas utilizam técnicas de perfilamento baseadas em dados para criar perfis digitais de indivíduos e grupos em uma ampla variedade de contextos, classificando indivíduos em categorias para auxiliar na tomada de decisões. Quando usada para automatizar e informar a tomada de decisões que afetam substancialmente os direitos e interesses significativos das pessoas, o perfilamento baseado em dados pode ter consequências graves (Yeung, 2018, p. 22-23).

Para o indivíduo afetado, a oportunidade de participar, contestar ou de outra forma desafiar o resultado da decisão e/ou o raciocínio subjacente sobre o qual a

decisão foi baseada, ou a qualidade ou integridade dos dados usados para informar a decisão, são praticamente inexistentes na prática. Embora o direito a um julgamento justo englobe uma série de direitos processuais mais específicos, incluindo o direito da pessoa de conhecer os motivos das decisões que as afetam adversa e significativamente, os sistemas usados para informar a tomada de decisões podem não ser configurados para produzir explicações significativas em termos inteligíveis para o indivíduo afetado (Yeung, 2018, p. 22-23).

Em razão da opacidade desses sistemas, de acordo com o Comitê, as preocupações podem ser ainda mais alarmantes, considerando a complexidade técnica, as dificuldades em avaliar a qualidade e a procedência dos dados e o fato do algoritmo desfrutar de proteção de propriedade intelectual como um segredo comercial. Consequentemente, esses sistemas correm o risco de interferir nos direitos ao devido processo legal, incluindo a presunção de inocência, especialmente em circunstâncias em que as consequências para o indivíduo afetado são graves e limitadoras da vida (Yeung, 2018, p. 22-23).

Especialmente preocupante é o uso crescente de sistemas de IA em contextos de justiça criminal para informar decisões de custódia e sentença, principalmente nos EUA, embora eles estejam sendo adotados em outros lugares (incluindo o Reino Unido) (Yeung, 2018, p. 22-23).

3.3.1.2 *O direito à liberdade de expressão:*

A operação do perfilamento algorítmico pode afetar significativamente o direito à liberdade de expressão, que inclui o direito de receber e transmitir informações, dada a influência poderosa que as plataformas digitais globais exercem sobre o ambiente informativo, tanto em nível individual quanto coletivo. Por exemplo, mecanismos de busca automatizados atuam como guardiões essenciais para pessoas que desejam buscar, receber ou transmitir informações, já que o conteúdo que não é indexado ou classificado em posições altas tem menos probabilidade de alcançar uma grande audiência (Yeung, 2018, p. 22-23).

No entanto, algoritmos de busca são intencionalmente projetados para servir aos interesses comerciais de seus proprietários e, portanto, inevitavelmente tendem a favorecer certos tipos de conteúdo ou provedores de conteúdo (Yeung, 2018, p. 24).

Geralmente, são algoritmos automatizados, e não seres humanos, que decidem como lidar, priorizar, distribuir e deletar conteúdo de terceiros em plataformas online, incluindo o tratamento de conteúdo durante campanhas políticas e eleitorais, por exemplo. Essas práticas não apenas implicam o direito individual à liberdade de expressão, mas também o objetivo de criar um ambiente propício para o debate público pluralista, igualmente acessível e inclusivo a todos (Yeung, 2018, p. 24).

De acordo com o Comitê, as plataformas online estão sob pressão crescente para combater ativamente discursos de ódio online por meio de técnicas automatizadas que detectam e removem conteúdo ilegal, especialmente após a transmissão ao vivo nas redes sociais do ataque a civis por um terrorista solitário em Christchurch no início de 2019. O uso generalizado de algoritmos para processos de filtragem e remoção de conteúdo, incluindo em plataformas de mídia social, também levanta preocupações, suscitando questões de legalidade, legitimidade e proporcionalidade (Yeung, 2018, p. 24).

Embora suas intenções sejam bem-vindas, falta transparência e responsabilidade em relação ao processo ou aos critérios adotados para determinar o que é considerado conteúdo "extremista" ou "claramente ilegal". Esses arranjos criam o risco de interferência excessiva no direito à liberdade de expressão e podem ser entendidos como uma transferência das responsabilidades de aplicação da lei dos Estados para empresas privadas (Yeung, 2018, p. 24).

O Comitê acrescenta ainda que regimes jurídicos que exigem que intermediários restrinjam o acesso a conteúdo com base em noções vagas como "extremismo", violam o princípio estabelecido de que intermediários não devem ser obrigados a conduzir monitoramento geral devido aos seus potenciais "efeitos inibidores" sobre a liberdade de expressão (Yeung, 2018, p. 25).

Além disso, surgem preocupações relacionadas ao processo em razão da capacidade das plataformas de decidir por si mesmas o que constitui conteúdo "extremista" e, portanto, sujeito a remoção. As ferramentas e medidas por meio das quais são tomadas decisões de identificação e remoção repousam sobre os provedores privados e, a menos que essas medidas não estejam sujeitas a uma supervisão estatal significativa e eficaz, correm o risco de ultrapassar limites legal e constitucionalmente prescritos (Yeung, 2018, p. 25).

Embora a necessidade de agir de forma decisiva contra a disseminação de mensagens de ódio e fake news seja incontestável, tais práticas levantam preocupações consideráveis relacionadas à legalidade das interferências na liberdade de expressão. O conteúdo extremista ou o material que incita a violência muitas vezes é difícil de identificar, mesmo para um humano treinado, devido à complexidade de desentrelaçar fatores como contexto cultural e humor (Yeung, 2018, p. 25).

Atualmente, os algoritmos não são capazes de detectar ironia ou análises críticas. A filtragem do discurso para eliminar conteúdo prejudicial por meio de algoritmos, portanto, enfrenta um alto risco de bloqueio excessivo e remoção de discursos que além de não serem inofensivos, ainda contribuem positivamente para o debate público. Por outro lado, a capacidade das plataformas de disseminar mensagens em tempo real e em escala global amplifica substancialmente o alcance, o escopo e, portanto, o impacto do discurso de ódio (Yeung, 2018, p. 25).

A adoção de abordagens automatizadas para a filtragem de conteúdo online destaca os desafios da responsabilidade que a crescente dependência de sistemas algorítmicos na vida contemporânea gera: embora eles ofereçam os benefícios de escala, velocidade e eficiência em relação à tomada de decisões humanas, necessitam de supervisão humana, que nem sempre é realizada da maneira adequada (Yeung, 2018, p. 25).

3.3.1.3 *Direito à privacidade e proteção de dados*

O direito à vida privada e os direitos à proteção de dados sempre existiram, entretanto, nos últimos anos o foco neles cresceu demasiadamente devido à capacidade dos algoritmos de facilmente coletar e reaproveitar vastas quantidades de dados, incluindo dados pessoais obtidos a partir da observação dos comportamentos dos usuários nas redes sociais. O uso desses dados pessoais e seu subsequente reaproveitamento ameaça o direito de uma pessoa à "autodeterminação informativa", especialmente dado que, até mesmo dados banais e inócuos podem ser mesclados com outros conjuntos de dados e explorados de maneiras que podem gerar informações que permitam inferir detalhes pessoais bastante íntimos com um nível muito alto de precisão (Yeung, 2018, p. 25).

Embora as leis gerais de proteção de dados mundialmente criadas nos últimos anos sejam uma salvaguarda importante, conferindo um conjunto de "direitos de proteção de dados" aos titulares, com o objetivo de protegê-los contra a coleta e o processamento desnecessários e ilegais de dados, eles podem não oferecer garantias abrangentes e eficazes na prática contra o uso alguns aplicativos (Yeung, 2018, p. 25).

3.3.1.4 *Proibição da discriminação no gozo de direitos e liberdades*

Como é sabido, a possibilidade de existência de viés e discriminação decorrentes do uso de técnicas de Machine Learning tem atraído considerável atenção, tanto de legisladores quanto de pesquisadores de IA. De acordo com o Comitê, existem muitas oportunidades para o viés afetar inadvertidamente os resultados produzidos por ferramentas de IA que utilizam Machine Learning e isso acontece principalmente em decorrência de: (i) preconceitos oriundos dos desenvolvedores dos algoritmos, (ii) preconceitos incorporados ao modelo com base no qual os sistemas são gerados, (iii) preconceitos inerentes aos conjuntos de dados usados para treinar os modelos ou (iv) preconceitos introduzidos quando esses sistemas são implementados em cenários do mundo real (Yeung, 2018, p. 26).

Não apenas sistemas de Machine Learning tendenciosos podem levar à discriminação e gerar decisões errôneas, mas isso pode envolver transgressões significativas, resultando em decisões sistematicamente tendenciosas contra grupos que historicamente foram socialmente desfavorecidos (e contra indivíduos que são membros desses grupos), reforçando assim e agravando a discriminação e a desvantagem estrutural, mesmo que esses efeitos não tenham sido pretendidos pelos programadores do sistema (Yeung, 2018, p. 26).

Ainda de acordo com o Comitê, essas preocupações têm sido particularmente agudas em relação ao uso de técnicas de aprendizado de máquina para decisões judiciais nos Estados Unidos, devido a alegações de que tais técnicas operam de maneira substancialmente tendenciosa contra negros e outras minorias. Em resposta a essas preocupações, tem surgido um crescente corpo de trabalho voltado para a criação de abordagens técnicas para combater esse viés (Yeung, 2018, p. 26).

Neste momento, após entender os riscos presentes no uso tanto de Machine Learning, quanto de inteligência artificial, surgem constantemente questionamentos acerca da responsabilidade pelas respostas e pelos atos praticados por esses softwares. Por isso, o próximo item analisará de maneira mais aprofundada essa temática.

3.4 Responsabilidade civil e o uso de inteligência artificial

Como demonstrado anteriormente, em alguns momentos, as novas tecnologias podem gerar sérias ameaças e riscos para os interesses e valores individuais e coletivos, podendo perpetuar a prática de infrações substanciais e sistemáticas, incluindo violações dos direitos humanos. Em conjunto, essas ameaças têm condições de comprometer ainda mais a saúde das bases morais e sociais coletivas das sociedades democráticas. Portanto, far-se-á necessário demonstrar e nomear de quem seria a responsabilidade com relação a prevenção, gestão e mitigação dessas ameaças, bem como pela reparação, caso se transformem em danos e violações de direitos, tanto para indivíduos quanto para grupos e à sociedade (Yeung, 2018, p. 37).

O Comitê, no artigo redigido, divide a discussão relacionada a responsabilidade em várias etapas. Em primeiro lugar, começa esclarecendo o que se entende por responsabilidade e por que ela é importante, enfatizando seu papel vital na garantia e expressão do Estado de Direito, essencial para a cooperação social pacífica (Yeung, 2018, p. 38).

Em segundo lugar, considera dois temas centrais levantados nas discussões contemporâneas sobre os riscos adversos associados às tecnologias de IA, especialmente o papel da indústria de tecnologia em promulgar e se comprometer voluntariamente a cumprir os chamados "padrões éticos" e o "problema de controle", que decorre da capacidade dos sistemas impulsionados por IA de operar de maneira autônoma, não mais necessitando ou dependendo de seus criadores humanos (Yeung, 2018, p. 38).

Em terceiro lugar, identifica uma variedade de "modelos de responsabilidade" que poderiam ser adotados para governar a alocação de responsabilidade de acordo com os diferentes tipos de impactos decorrentes da operação de sistemas de IA, o que inclui modelos baseados em intenção/culpabilidade, criação de

risco/negligência, responsabilidade estrita e esquemas de seguro obrigatório (Yeung, 2018, p. 38).

Em quarto lugar, chama a atenção para os desafios na alocação de responsabilidade gerados pela operação de sistemas sócio-técnicos complexos e interativos, que envolvem contribuições de múltiplos indivíduos, organizações, componentes de máquinas, algoritmos de software e usuários humanos, frequentemente em ambientes complexos e altamente dinâmicos (Yeung, 2018, p. 38).

Em quinto lugar, destaca uma variedade de mecanismos não judiciais para assegurar tanto a responsabilidade prospectiva quanto histórica pelos impactos dos sistemas de IA, de modo a incluir vários tipos de avaliações, técnicas de auditoria e mecanismos de proteção técnica (Yeung, 2018, p. 38).

Em sexto lugar, enfatiza o papel e as obrigações dos Estados em relação aos riscos associados às tecnologias digitais avançadas, concentrando-se especificamente em suas obrigações de garantir a proteção efetiva dos direitos humanos (Yeung, 2018, p. 38).

Por fim, o Comitê destaca a necessidade de revitalizar o discurso sobre direitos humanos em uma era digital, chamando a atenção para a importância de proteger e cultivar as bases sócio-técnicas necessárias. De acordo com ela, sem tais bases, os direitos e liberdades humanas não podem ser exercidos de maneira prática ou significativa (Yeung, 2018, p. 38).

3.4.1 Responsabilidade de acordo com a lei

De acordo com o Comitê, a responsabilidade básica é essencial, não apenas para a autocompreensão das pessoas como indivíduos autores de suas próprias vidas, mas também como membros de uma comunidade de agentes morais. Agentes morais têm a capacidade e liberdade para fazer escolhas sobre suas ações, podendo fazê-lo de maneiras que sejam injustas ou causem danos, seja a outros indivíduos ou às condições essenciais para manter a estabilidade e a cooperação sociais necessárias para sustentar a vida comunitária (Yeung, 2018, p. 40).

Em uma comunidade, é normal e até mesmo necessário que os seus membros responsabilizem uns aos outros por seus atos, e ao final, isso é exatamente o que caracteriza uma comunidade política como uma comunidade

moral (ou seja, uma comunidade de agentes morais). Nesse momento, o respeito mútuo e o autocontrole dos membros de uma comunidade moral, que são a base de um Estado de Direito, é o que torna sustenta e torna possível a vida comunitária (Yeung, 2018, p. 40).

Desta maneira, caso uma sociedade careça de um sistema capaz de institucionalizar as práticas de responsabilização dos indivíduos, analisando e eventualmente punindo-os pelos danos que suas condutas causem a terceiros, ela terminaria extinguindo-se, posto que a ausência de reação a atitudes reprováveis prejudica a cooperação social e pacífica entre as pessoas, desmantelando, consequentemente, o Estado de Direito (Yeung, 2018, p. 40).

Em outras palavras, nosso sistema, para garantir que a responsabilidade seja devidamente alocada, desempenha um papel crítico na sustentação da estrutura social subjacente à cooperação, sem a qual a lei não pode governar. Ao mesmo tempo, é importante reconhecer que a estabilidade e continuidade dessas bases sociais repousam, em última análise, no respeito mútuo e autocontrole dos membros individuais da comunidade moral e não em um sistema de coerção e controle tecnológico (Yeung, 2018, p. 40).

3.4.2 Responsabilidade, responsabilização e transparência

Como explicado anteriormente, o ponto nevrálgico do Estado de Direito está na sua capacidade de criar sistemas competentes para garantir que as comunidades moral e política estejam comprometidas com o respeito aos direitos humanos, estabelecendo e implementando mecanismos institucionais para responsabilizar os membros da comunidade por suas condutas. Embora o conceito de responsabilidade seja contestado, para os propósitos presentes, ele foi útilmente descrito como "exigir que uma pessoa explique e justifique - em relação a algum critério - suas decisões ou atos e, em seguida, reparar qualquer falha ou erro" (Oliver, 1994, p. 245).

Para o direito brasileiro, a responsabilidade encontra-se descrita no título IX do Código Civil, principalmente no artigo 927, que determina que “aquele que, por ato ilícito, causar dano a outrem, fica obrigado a repará-lo”. Desta maneira, é possível dizer que o sistema jurídico brasileiro especifica e espera que haja um

compromisso com a reparação das falhas por aquele que a causou (Yeung, 2018, p. 41).

Portanto, de acordo com o entendimento do Comitê, os mecanismos de responsabilidade possuem as seguintes quatro características: (i) obter as justificativas dos indíviduos para as suas ações; (ii) estabelecer padrões para julgar essas justificativas; (iii) julgar essas justificativas e; (iv) decidir quais consequências (se houver) devem ser aplicadas. Em última instância, o que se busca afirmar é que o conceito de responsabilidade é de particular importância nas relações entre o Estado e seus indivíduos, em que se espera que os indivíduos ajam em busca de um Estado democrático, prestando conta e sendo responsável por aquilo que contrarie a conduta moral esperada (Yeung, 2018, p. 41).

A transparência está diretamente ligada à responsabilidade, na medida em que a responsabilidade exige que aqueles chamados a prestar contas possam explicar as razões de suas ações e justificar essas ações de acordo com um conjunto específico de regras ou padrões de avaliação. Portanto, a transparência é importante para que a parte afetada avalie a qualidade dessas razões (Yeung, 2018, p. 41).

Mecanismos de responsabilidade têm importância particular em relação ao exercício do poder governamental em sociedades democráticas liberais, porque os funcionários governamentais são considerados servos dos cidadãos em nome dos quais agem e de quem seu poder é, em última análise, derivado. No entanto, a importância da responsabilidade surge sempre que o exercício do poder tem a capacidade de afetar outras pessoas de maneiras adversas (Yeung, 2018, p. 41).

Destarte, o maior aprofundamento de estudos sobre o poder, escalabilidade e efeitos de sistemas complexos que dependem de tecnologias de IA deram origem a um conjunto de preocupações que podem ser compreendidas como unidas por uma preocupação em garantir a "responsabilização algorítmica", especialmente dada a opacidade desses sistemas e sua potencial utilização de maneiras que podem ter implicações altamente danosas para indivíduos, grupos e a sociedade em geral. Garantir a responsabilização por violações de direitos humanos e outras consequências adversas resultantes da operação dessas tecnologias é, portanto, essencial (Yeung, 2018, p. 41).

Considerando o entendimento do Comitê, embora as leis existentes, incluindo a lei de proteção de dados, o código de defesa do consumidor, a lei de concorrência

e a constituição, que consagram a proteção dos direitos humanos nos sistemas legais nacionais, tenham o potencial de desempenhar um papel significativo e importante na garantia de várias dimensões da responsabilidade algorítmica, sua contribuição para garantir essa responsabilidade está além do escopo deste estudo. Pelo contrário, a discussão a seguir procura examinar as implicações das tecnologias digitais avançadas (incluindo sistemas de IA) para o conceito de responsabilidade, concentrando-se principalmente em suas implicações para violações de direitos humanos, recorrendo tanto à filosofia moral quanto à pesquisa jurídica (Yeung, 2018, p. 41).

3.5 Dimensões da responsabilidade

Este conceito geral de responsabilidade como 'prestar contas' foi extensivamente examinado na literatura jurídica e filosófica, e vários insights dessa literatura são selecionadamente utilizados na análise realizada pelo Comitê.

Embora haja muitos sentidos diferentes nos quais o termo responsabilidade é utilizado, para o Comitê, o elemento temporal da responsabilidade é digno de ênfase, voltando-se em duas direções (Hart, 2008, p. 230):

(a) Responsabilidade histórica (ou retrospectiva): que olha para trás, buscando alocar responsabilidade por condutas e eventos que ocorreram no passado. Como veremos, consideráveis dificuldades são encontradas ao alocar responsabilidade histórica por danos e injustiças causados por sistemas de IA (Yeung, 2018, p. 42).

(b) Responsabilidades prospectivas: que estabelecem obrigações e deveres associados a papéis e tarefas que visam o futuro, direcionados para a produção de resultados positivos e a prevenção de resultados negativos. Responsabilidades prospectivas servem a uma função orientadora importante (Yeung, 2018, p. 42).

Como Cane (2002, p. 45) coloca, “uma das razões mais importantes pelas quais nos interessamos por responsabilidade e conceitos relacionados é por causa do papel que desempenham no raciocínio prático sobre nossos direitos e obrigações em relação a outras pessoas, e sobre a maneira como devemos nos comportar em nossos tratos com elas”. No contexto da responsabilidade pelas ações e

consequências resultantes de sistemas autônomos de IA, a ideia de papel e responsabilidade (Hart, 2008, p. 211-230) às vezes foi destacada (Yeung, 2018, p. 42).

Qualquer resposta legítima e eficaz às ameaças, riscos, danos e violações de direitos apresentados por tecnologias digitais avançadas provavelmente exigirá um foco nas consequências para indivíduos e sociedade que atenda, e possa garantir que, tanto a responsabilidade prospectiva voltada para prevenir e mitigar riscos, quanto a responsabilidade histórica por efeitos adversos decorrentes da operação de sistemas sócio-técnicos complexos nos quais essas tecnologias estão incorporadas, sejam devidamente e justamente atribuídas (Yeung, 2018, p. 42).

Apenas se ambas as dimensões históricas e prospectivas da responsabilidade forem consideradas é que indivíduos e a sociedade podem ter confiança de que esforços serão feitos primeiro, para evitar danos e injustiças de ocorrerem, e segundo, se ocorrerem, então mecanismos institucionais podem ser confiáveis para garantir uma reparação apropriada, e correção capaz de evitar danos ou injustiças adicionais. Isso exigirá um foco tanto naqueles envolvidos no desenvolvimento e implementação dessas tecnologias, quanto nos usuários individuais, nos grupos afetados por elas, e na ação do Estado para garantir o estabelecimento e a manutenção das condições necessárias para proteger os cidadãos contra ameaças e riscos inaceitáveis, garantindo assim que os direitos humanos sejam adequadamente protegidos (Yeung, 2018, p. 42).

Em outras palavras, uma abordagem adequada da responsabilidade das tecnologias e sistemas de IA atenderá às posições tanto do agente moral quanto do paciente moral, bem como à comunidade moral mais ampla em geral (Liu; Zawieska, 2020).

Após esclarecer o que se entende por responsabilidade e destacar a necessidade de considerar tanto suas dimensões prospectivas quanto retrospectivas, far-se-á relevante adentrar onde reside a responsabilidade pelas consequências adversas, ameaças e riscos associados ao desenvolvimento e implementação de tecnologias de IA, incluindo violações de direitos humanos e outros erros e danos decorrentes de sua operação. Como observou o Grupo Europeu de Ética da UE, as tecnologias de IA levantam "questões sobre a responsabilidade moral humana. Como a responsabilidade moral deve ser atribuída e distribuída, e quem é responsável (e em que sentido)?" (Yeung, 2018, p. 43).

Em outras palavras, a complexidade das próprias tecnologias e dos contextos sócio-técnicos mais amplos nos quais são implementadas e aplicadas pode obscurecer as linhas de responsabilidade moral, especialmente quando operam de maneiras inesperadas gerando danos ou violando direitos. No entanto, devemos ter em mente que responsabilidade moral e responsabilidade legal são conceitos distintos, embora relacionados (Yeung, 2018, p. 43).

Ao contrário da moralidade, a lei possui um sistema altamente desenvolvido para institucionalizar e aplicar a responsabilidade (incluindo a aplicação de sanções em determinadas circunstâncias), porque precisa julgar disputas do mundo real, o que requer tanto a finalidade do julgamento quanto a certeza legal (Cane, 2002).

Uma sociedade não pode depender exclusivamente das inclinações individuais para 'agir eticamente' porque a falta de mecanismos institucionais para impor esses padrões (incluindo autoridade legal para sancionar a não conformidade) significaria que o sistema seria inteiramente voluntário, correndo claro risco de falhar em fornecer as bases sociais estáveis e confiáveis necessárias para uma cooperação social confiável e pacífica dentro das sociedades contemporâneas. Por esse motivo, o papel da lei de assegurar e institucionalizar a responsabilidade de modo a garantir a proteção dos direitos e fazer cumprir o desempenho dos deveres legais é essencial (Yeung, 2018, p. 42).

Como a discussão a seguir demonstra, a maneira como os sistemas legais têm alocado a responsabilidade histórica geralmente é mais sensível aos interesses das vítimas e da sociedade na segurança da pessoa e da propriedade em comparação com as descrições filosóficas morais da responsabilidade, que tendem a se concentrar na conduta do agente moral e se ela atrai adequadamente a culpa. No entanto, aplicar esses conceitos morais e legais de responsabilidade ao desenvolvimento e implementação de tecnologias digitais avançadas (incluindo IA) em contextos contemporâneos pode não ser algo fácil e que possa ser realizado diretamente (Yeung, 2018, p. 42).

A capacidade dessas tecnologias de operar de maneiras que antes não eram possíveis pode desafiar as concepções tradicionais de responsabilidade civil, moral e social, especialmente dadas as propriedades anteriormente identificadas como relevantes para a responsabilidade, incluindo:

(a) Inscrutabilidade e opacidade;

(b) Natureza complexa e dinâmica;
(c) Dependência de entrada, interação e discrição humanas;
(d) Natureza de propósito geral;
(e) Interconectividade, escalabilidade e ubiquidade global;
(f) Operação automatizada e contínua, muitas vezes em tempo real;
(g) Dependência de grandes conjuntos de dados;
(h) Capacidade de gerar insight 'oculto' a partir da fusão de conjuntos de dados;
(i) Capacidade de imitar com precisão traços humanos;
(j) Maior complexidade de software (incluindo vulnerabilidade a falhas e ataques maliciosos);
(k) Capacidade de 'personalizar' e configurar ambientes de escolha individuais;
(l) Capacidade de redistribuir riscos, benefícios e ônus entre indivíduos e grupos por meio do uso de sistemas de otimização impulsionados por IA que reconfiguram ambientes sociais e arquiteturas de escolha;
(m) Capacidade de gerar problemas de ação coletiva.

Antes de prosseguir, é importante esclarecer a distinção conceitual entre dois tipos diferentes de efeitos adversos que podem (e têm) surgido da operação de sistemas de IA:

(a) Violações de direitos humanos, incluindo, mas não se limitando aos direitos protegidos pela Convenção Europeia de Direitos Humanos (ECHR);
(b) Danos tangíveis à saúde humana, propriedade ou meio ambiente.

Esses são conceitos e consequências separados e distintos. É possível que ocorra uma violação dos direitos humanos sem nenhum dano tangível, e vice-versa. Por exemplo, a remoção em 2016 pelo Facebook da icônica fotografia de uma menina de 9 anos nua fugindo de bombas durante a Guerra do Vietnã, sob o argumento de que a nudez violava suas normas comunitárias, pode ser entendida como uma violação do direito à liberdade de expressão e informação do Artigo 10 da

Convenção Europeia de Direitos Humanos, embora não tenha gerado nenhum dano tangível substancial (Scott; Isaac, 2016).

Por outro lado, se um carro autônomo colidir com e ferir um animal selvagem, configuraria-se um dano sem violação dos direitos humanos. No entanto, qualquer evento ou série de eventos pode envolver tanto danos tangíveis quanto uma violação dos direitos humanos. Assim, se um veículo autônomo colidir com e fatalmente ferir um pedestre, isso implicaria tanto uma violação do direito à vida do Artigo 2, ao passo que causaria um dano tangível (Yeung, 2018, p. 44).

Nesse momento, o foco deste estudo está em analisar as implicações da responsabilidade dos sistemas de IA sob uma perspectiva de direitos humanos. Portanto, aprofundará com a explanação das responsabilidades por violações de direitos humanos, em vez da responsabilidade por danos tangíveis decorrentes da operação desses sistemas.

A discussão a seguir se concentra principalmente naqueles que criam, desenvolvem, implementam e supervisionam sistemas de IA. Pergunta-se se eles podem ser responsabilizados pelas consequências adversas que esses sistemas podem gerar, começando com um exame de dois temas centrais que surgiram em respostas recentes, onde havia a intenção de identificar onde reside a responsabilidade pelos riscos que as tecnologias de IA podem representar: primeiro, ação voluntária da indústria de tecnologia ao promulgar e proclamar publicamente seu compromisso com as chamadas "diretrizes éticas", e segundo, alegações de que, como os sistemas de IA agem autonomamente, isso isenta seus criadores da responsabilidade por suas decisões e quaisquer efeitos adversos consequentes.

As obrigações do Estado em relação a esses efeitos adversos são consideradas após a descrição de vários "modelos de responsabilidade" que podem ser aplicados em situações envolvendo àqueles que desenvolvem e implementam sistemas de IA.

3.5.1 Códigos de Ética e o projeto IA responsável

De acordo com o Comitê, o aumento da ansiedade pública e o recente *“Techlash”* em resposta às crescentes práticas e políticas das grandes empresas de tecnologia, especialmente após o uso de microdirecionamento político e o escândalo da Cambridge Analytics, precipitaram numerosas iniciativas voluntárias de “ética”

pela indústria de tecnologia. Essas iniciativas geralmente envolvem a promulgação de um conjunto de normas e padrões, seja por empresas individuais de tecnologia ou por um grupo delas, que publicamente e voluntariamente se comprometem a cumprir esses padrões de conduta divulgados. Essas iniciativas podem ser entendidas como parte de um movimento em direção ao que Liu e Zawieska chamam de projeto 'IA/robótica responsável' (Liu; Zawieska, 2017).

Dois aspectos dessas iniciativas valem a pena destacar. Em primeiro lugar, elas estão preocupadas com a responsabilidade prospectiva, buscando identificar e alocar esferas de obrigação para aqueles envolvidos em cada estágio do design, desenvolvimento e implementação dessas tecnologias, com o objetivo de demonstrar ao público a seriedade de seu compromisso em lidar com preocupações éticas (Liu; Zawieska, 2017).

Uma característica notável dessas iniciativas é que elas tendem a evitar explicitamente fazer referência às responsabilidades históricas daqueles envolvidos no design, desenvolvimento e implementação dessas tecnologias quando as coisas dão errado. Elas também não costumam especificar sobre quem deve recair a culpa por tais consequências, nem reconhecem qualquer obrigação de compensar aqueles afetados adversamente.

Ao invés disso, como Liu explica, nessa situação há um senso de responsabilidade que se atribui a um indivíduo em virtude do cargo que ocupa ou da função que se espera que ele cumpra, e, portanto, é pela execução de obrigações conectadas ao papel de um indivíduo e que podem ser pré-definidas e especificadas antecipadamente que se constrói o caminho para essa tentativa de ética (Cane, 2002, p. 32). Assim, caso o indivíduo tenha cumprido os deveres associados ao seu papel ou cargo, considera-se que suas responsabilidades foram também cumpridas, numa tentativa de limitar a responsabilidade daqueles que seriam os principais pontos focais de um sistema de IA.

Em segundo lugar, essas iniciativas de 'IA/robótica responsável' podem ser caracterizadas como um movimento emergente de autorregulação profissional que pode ser situado dentro de um fenômeno social mais antigo frequentemente discutido sob o título de responsabilidade social corporativa. A natureza desses chamados códigos éticos como sociais, em vez de legais, e sua aparente voluntariedade permitem que se chegue a conclusão de que as obrigações e

compromissos especificados nesses códigos não são legalmente exigíveis se violadas (Yeung, 2018, p. 46).

Segundo o Comitê, essas iniciativas também não costumam prever o estabelecimento de mecanismos de auditoria, através dos quais um órgão independente e externo seja capacitado para avaliar até que ponto esses compromissos foram cumpridos ou impor sanções por não conformidade. Assim, embora essas iniciativas reconheçam positivamente que o desenvolvimento ético de tecnologias digitais avançadas é uma questão de preocupação pública que merece sua ação e atenção, essas iniciativas carecem de mecanismos institucionais formais para fazer cumprir e sancionar violações (Yeung, 2018, p. 46).

Outro possível ponto de crítica destacado pelo Comitê é o fato de também não haver representação sistemática do público na definição desses padrões, o que fez com que essas iniciativas fossem amplamente reprovadas e consideradas como uma forma de lavagem ética, que falha em levar a sério as suas próprias preocupações criticadas como uma forma de “lavagem ética” (Wagner, 2019; Metzinger 2019), falhando em levar a sério as preocupações éticas (Greene; Hoffmann; Stark, 2019; Hagendorf, 2019).

Se esses códigos fossem apoiados por mecanismos institucionais, respaldados por lei, incluindo disposições para participação externa na definição e avaliação dos padrões em si, e supervisão independente para avaliar se empresas e organizações individuais cumpriram de fato as normas e padrões especificados, haveria uma base mais forte na qual os afetados (e a sociedade em geral) poderiam ter confiança de que salvaguardas significativas e democraticamente legítimas estão em vigor para prevenir e mitigar alguns dos riscos éticos associados a essas tecnologias (Nemitz, 2018).

Desta maneira, pela perspectiva dos direitos humanos, a necessidade de salvaguardas significativas e eficazes é um dos pontos principais dessa discussão. Ao mesmo tempo, abordagens prospectivas não podem garantir que a responsabilidade histórica, no caso de ocorrer dano ou má conduta, será devidamente atribuída. Como argumentam Liu e Zaweiska, embora o projeto “IA responsável” possa ser bem-vindo, deixa uma lacuna de responsabilidade, porque está preocupado apenas com a responsabilidade profissional, e não com a responsabilidade em sentido mais amplo, que a Comissão denomina como “responsabilidade causal” (Yeung, 2018, p. 47).

Ao contrário da responsabilidade profissional, a responsabilidade causal é uma forma de responsabilidade histórica. Sua preocupação é identificar e estabelecer uma relação entre causa e efeito. É, portanto, retrospectiva por natureza, inerentemente voltada para fora e orientada para a relação, porque destaca o paciente moral (ou seja, a pessoa ou pessoas prejudicadas pela atividade relevante) (Yeung, 2018, p. 47).

Em resumo, a responsabilidade profissional, que se concentra em atribuir incumbências e deveres aos indivíduos de acordo com seus cargos e funções, apesar de ser relevante, resta não sendo suficiente. Claramente, existe uma lacuna já que a responsabiliadade somente profissional cumpre funções somente prospectivas e não garante que aquele que eventulmente seja prejudicado receba uma compensação, contrariando o que prevê o ordenamento civil e constitucional brasileiro.

Em outras palavras, a designação de responsabilidade profissional não pode garantir a responsabilidade retrospectiva nem atribuir culpa, porque está preocupada apenas com o cumprimento de obrigações pré-estabelecidas e não a responsabilização considerando as consequências que as ações geraram (Yeung, 2018, p. 47).

3.5.2 A autonomia da IA e o desafio em controlá-la

Uma alegação frequente em resposta às preocupações sobre a necessidade de identificar onde reside a responsabilidade pelas implicações adversas das tecnologias digitais avançadas é que, em razão desses sistemas operarem autonomamente e, portanto, sem intervenção e controle humano, aqueles que os desenvolvem e implementam não podem ser considerados responsáveis por suas decisões, ações e consequências correspondentes (Yeung, 2018, p. 47).

Essa visão foi delineada por Matthias (2004, p. 175 *apud* Yeung, 2018, p. 47), que argumenta que o agente só pode ser considerado responsável se ele conhecer os fatos particulares que cercam sua ação e se for capaz de formar livremente uma decisão de agir e selecionar uma das ações alternativas disponíveis com base nesses fatos.

Uma classe crescente de máquinas, que Matthias se refere como “agentes artificiais autônomos” (2004, p. 175 *apud* Yeung, 2018, p. 47), é capaz de cumprir

alguns propósitos, muitas vezes bastante estreitos, movendo-se autonomamente e agindo sem supervisão humana. Esse agente pode ser um programa de software que se move por um espaço de informação, mas também pode ter uma presença física e se mover no tempo e no espaço (Yeung, 2018, p. 47).

Esses agentes são deliberadamente projetados para agir e, inevitavelmente, interagir, com outras coisas, pessoas e entidades sociais (leis, instituições e expectativas). Pelo menos para aqueles que têm uma presença física e podem aprender com a interação direta em ambientes reais, eles podem, com essa troca, manipular diretamente esse mesmo ambiente e compartilhar seu ambiente com os humanos (Yeung, 2018, p. 47).

Matthias (2004, p. 175 *apud* Yeung, 2018, p. 47) argumenta que surge uma "lacuna de responsabilidade" porque, em situações como as descritas anteriormente, o agente humano que o programou não exerce mais controle direto sobre o comportamento da máquina, que passa a, gradualmente, transferido para a máquina em si. Por conseguinte, seria injusto responsabilizar os humanos por ações de máquinas sobre as quais não poderiam ter controle suficiente. Ele oferece vários exemplos desses tipos de agentes de máquinas, incluindo aqueles que dependem de:

(a) A operação de redes neurais artificiais: em vez de uma representação clara e distinta de informações e controle de fluxo, tem-se uma matriz às vezes muito grande de pesos sinápticos, que não podem ser interpretados diretamente. Assim, o conhecimento e o comportamento armazenados em uma rede neural só podem ser inferidos indiretamente por meio de experimentação e da aplicação de padrões de teste após o treinamento da rede ser concluído (Matthias, 2004, p. 175 *apud* Yeung, 2018, p. 47);

(b) Aprendizado por reforço: geralmente baseado nos mesmos conceitos de redes neurais, mas adicionalmente ele elimina a distinção entre uma fase de treinamento e uma fase de produção. Sistemas de aprendizado por reforço exploram seu espaço de ação enquanto trabalham em seu ambiente operacional, que é sua característica central (permitindo-lhes se adaptar a ambientes sempre em mudança), sendo essa uma grande desvantagem em relação à previsibilidade. A informação armazenada

> na rede não pode ser totalmente verificada, nem mesmo indiretamente, porque ela está sempre mudando. Mesmo provando-se matematicamente que o desempenho geral de tal sistema eventualmente convergirá, haverá erros inevitáveis no caminho para esse estado otimizado. O criador de tal sistema (que Matthias comenta que não é realmente um programador no sentido tradicional) não pode eliminar esses erros, pois eles devem ser explicitamente permitidos para que o sistema permaneça operacional e melhore seu desempenho (Matthias, 2004, p. 175 *apud* Yeung, 2018, p. 47);
>
> (c) Métodos de programação genética nos quais uma camada adicional de código gerado por máquina opera entre o programador e o produto da programação. Ao contrário das redes neurais, onde o designer ainda define os parâmetros operacionais do sistema (a arquitetura da rede, as camadas de entrada e saída e sua interpretação) e pelo menos define o alfabeto usado e a semântica dos símbolos, nesse caso, o programador genético perde até mesmo essa quantidade mínima de controle, pois cria-se uma máquina que se programa (Matthias, 2004, p. 175 *apud* Yeung, 2018, p. 47).

Ao mesmo tempo, Matthias (2004 *apud* Yeung, 2018, p. 47) observa que agentes autônomos privam o programador de um vínculo espacial entre o programador e o agente da máquina resultante. Portanto, o agente da máquina age fora do horizonte de observação do programador e pode não ser capaz de intervir manualmente (no caso de uma falha ou erro, que pode ocorrer em um ponto muito posterior no tempo).

Assim, esses processos envolvem o designer de máquinas, perdendo cada vez mais o controle sobre elas, transferindo gradualmente o controle para a máquina em si, e nesse momento, de acordo com Matthias, o papel do programador muda de codificador para criador de organismos de software. À medida que a influência do criador da máquina diminui, a influência do ambiente operacional aumenta, de modo que o programador transfere seu controle sobre o produto para o ambiente, especialmente no caso das máquinas que continuam a aprender e se adaptar em seu ambiente operacional final (Matthias, 2004, p. 175 *apud* Yeung, 2018, p. 47).

Dado que esses agentes terão que interagir com uma variedade e número potencialmente grandes de pessoas (usuários) e situações, geralmente não será possível para o criador prever ou controlar a influência do ambiente operacional. Segundo Matthias, o resultado líquido é que essas máquinas operam além do controle de seus criadores e, portanto, podem causar danos pelos quais não se pode justamente responsabilizá-los. No entanto, Matthias argumenta que, como não se pode abdicar de usar tais sistemas, deve-se encontrar uma maneira de solucionar a lacuna de responsabilidade considerando a prática moral e legislação (Matthias, 2004, p. 183 *apud* Yeung, 2018, p. 48).

3.5.3 Teorias de responsabilidade moral baseadas em escolhas

A afirmação de Matthias (2004, p. 183 *apud* Yeung, 2018, p. 48) de que aqueles que criam máquinas autônomas não podem ser responsabilizados por suas ações se baseia em uma explicação de responsabilidade moral baseada em escolhas que tem tendido a dominar a reflexão acadêmica contemporânea sobre as implicações éticas e morais da IA. De acordo com o entendimento que considera a responsabilidade moral como sendo baseada em escolhas, a conduta relaciona-se diretamente com a ação quando existe dano, tendo sido a conduta ilícita pretendida pelo indivíduo (Wallace, 1994).

Nessa explicação, um agente (X) somente é moralmente responsável por um resultado (Y) se X "causou" Y. Para estabelecer que X causou Y, então X deve ter se envolvido em conduta pela qual X pode ser considerado causalmente responsável.

Estabelecer esse elo causal requer que X tenha escolhido voluntariamente se envolver na conduta, mesmo que essa conduta acabe tendo consequências e efeitos que X não pretendia ou queria. Segundo Matthias, considerando que os sistemas de IA têm a capacidade de tomar suas próprias decisões de maneiras que não foram pré-programadas por desenvolvedores humanos, os desenvolvedores não têm o grau necessário de controle e, portanto, não são moralmente responsáveis pelas decisões desses agentes computacionais ou suas consequências (Matthias, 2004, p. 183 *apud* Yeung, 2018)

De acordo com o Comitê, a validade da afirmação de que a capacidade de os agentes computacionais agirem autonomamente quebra a cadeia de causalidade entre os atos de seus desenvolvedores e as decisões tomadas por esses agentes é

altamente discutível. Como questão preliminar, é importante reconhecer que teorias de responsabilidade moral baseadas em escolhas são particularmente inadequadas como modelo para identificar responsabilidade por violações de direitos humanos (Yeung, 2018, p. 50).

É inerente à natureza e ao conceito de direitos em geral, e direitos humanos em particular, que eles protegem valores de importância fundamental, de modo que qualquer interferência com eles atrai responsabilidade per se, sem a necessidade de prova de culpa.

O artigo do Comitê traz um exemplo relevante para essa situação, que seria: considere-se novamente o exemplo da remoção pela Facebook da icônica imagem da menina vietnamita em 2016. Em circunstâncias em que a legislação nacional impõe obrigações legais tanto a atores estatais quanto não estatais de respeitar os direitos humanos, o Facebook seria considerado legalmente responsável por violar o direito à liberdade de expressão, sem a necessidade de demonstrar que tinha a capacidade de controlar se a imagem seria removida.

Em outras palavras, uma violação do direito à liberdade de expressão ocorreu mesmo se a decisão de a remover tivesse sido tomada por um sistema algorítmico automatizado agindo independentemente, sem intervenção humana direta, e mesmo que os designers humanos do sistema automatizado não tivessem a intenção ou previsão de que a imagem específica em questão poderia ser removida automaticamente (Yeung, 2018, p. 50).

3.6 Alocação de responsabilidades

Embora o modelo de responsabilidade que se aplica a violações de direitos humanos seja amplamente entendido como um modelo de responsabilidade estrita, sem a necessidade de prova de culpa, a alocação de obrigações de reparo por danos tangíveis à saúde ou propriedade pode ser legalmente distribuída de acordo com uma variedade de modelos de responsabilidade. Como visto anteriormente, os sistemas de IA podem operar de maneiras que resultam tanto em violações de direitos humanos quanto em danos a indivíduos e/ou propriedades, e a alocação de responsabilidade histórica por danos serve como uma orientação para aqueles envolvidos no design, desenvolvimento, produção e implementação de sistemas de IA, de modo a especificar a natureza e o alcance de suas obrigações.

De acordo com o Comitê, a variedade de modelos legais que podem ser aplicados para alocar e distribuir os danos decorrentes de uma conduta demonstra claramente que seria um equívoco esperar que um único modelo de responsabilidade se aplicaria de maneira justa a todos os diferentes tipos de consequências adversas que podem resultar do uso de tecnologias digitais avançadas.

Como observado anteriormente, ao contrário da análise filosófica de responsabilidade, que tende a focar nos agentes em detrimento das vítimas e da sociedade, os modelos legais de responsabilidade (Cane, 2002, p. 2) são relacionais no sentido de que estão preocupados não apenas com a posição de indivíduos cuja conduta atrai responsabilidade, ou seja, agentes morais, mas também com o impacto dessa conduta em outros indivíduos e na sociedade em geral (Cane, 2002, p. 4), conforme Peter Cane observou:

> A responsabilidade não é apenas uma função da qualidade de vontade manifestada na conduta, nem da qualidade dessa conduta. Também diz respeito ao interesse que todos compartilhamos na segurança da pessoa e da propriedade, e à maneira como recursos e riscos são distribuídos na sociedade. A responsabilidade é um fenômeno relacional (Cane, 2002, p. 109).

Em outras palavras, a responsabilidade legal enfatiza a relação entre agentes morais, pacientes morais e a sociedade em geral, em vez de se concentrar exclusivamente na conduta dos agentes morais e se essa conduta atrai de maneira justa a responsabilidade. Portanto, a análise acadêmica das várias maneiras como os sistemas legais nacionais alocam responsabilidade por condutas que causam danos ou outros eventos adversos demonstra como cada um desses modelos envolve um equilíbrio diferente de interesses entre agentes morais e pacientes morais (ou 'vítimas', como são comumente chamadas na pesquisa jurídica) (Yeung, 2018, p. 50).

No entanto, esta discussão não procura avaliar se as abordagens legais atuais adotadas dentro dos sistemas legais nacionais alocam adequadamente a responsabilidade por danos por meio da aplicação de regras nacionais de responsabilidade civil, especialmente porque a capacidade da lei nacional de alocar responsabilidade histórica por danos e erros causados por sistemas de IA ainda não foi totalmente testada por meio de litígios. Em vez disso, a discussão a seguir

delineia brevemente quatro modelos amplos de responsabilidade refletidos nos sistemas legais anglo-americanos, notadamente:

(1) Modelos baseados em intenção/culpabilidade;
(2) Modelos baseados em risco/negligência;
(3) Responsabilidade estrita; e
(4) Esquemas de seguro obrigatório.

Os modelos de responsabilidades à priori destacados tem como estratégia destacar a gama de modelos potenciais de responsabilidade que podem ser usados para alocar e distribuir ameaças, riscos e danos associados ao uso de tecnologias digitais avançadas (Cane, 2002). Esses esboços descrevem seletivamente o que o Comitê chama de “condição de controle/conduta” e a “condição epistêmica”, aplicáveis a cada modelo, em vez de fornecer uma descrição completa e detalhada do conteúdo e contornos de cada modelo.

Juntos, eles revelam como cada modelo atinge um equilíbrio diferente entre o interesse do indivíduo, como agentes, na liberdade de ação e o interesse do indivíduo, como vítima, levando em consideração os direitos pela perspectiva da segurança da pessoa e da propriedade (Cane, 2002, p. 98).

Isso sugere que identificar qual (se houver) desses modelos é mais apropriado para alocar e distribuir os vários riscos associados à operação de tecnologias digitais avançadas está longe de ser evidente, mas implicará uma escolha de política social sobre como esses ônus devem ser apropriadamente alocados e distribuídos (Danaher, 2016, p. 299).

3.6.1 Modelos baseados na culpabilidade

Os modelos baseados na culpabilidade, que na maioria das vezes é utilizado como principal modelo de responsabilidade para o direito penal, concentram-se principalmente na voluntariedade da conduta do agente. Eles podem ser interpretados como exigindo a satisfação de duas condições: em primeiro lugar, a

condição de controle, demonstrando que o agente era causalmente responsável pela conduta legalmente proscrita na medida em que o agente tinha uma escolha livre e voluntária sobre agir desta maneira, e em segundo lugar, a condição epistêmica, exigindo prova de culpa, amplamente entendida como exigindo que o agente tivesse conhecimento real e consciência dos fatos particulares que cercam as consequências prejudiciais da conduta do agente, e a ação do agente pode ser entendida como baseada nesses fatos (Cane, 2002, p. 79).

É um modelo de responsabilidade baseado em intenção/culpabilidade que sustenta as discussões sobre a responsabilidade moral que predominaram em debates orientados filosoficamente sobre se os desenvolvedores humanos de softwares que utilizam inteligência artificial são moralmente responsáveis pelas ações desses agentes.

Pelo menos por enquanto, porque os agentes computacionais carecem da capacidade de conhecimento subjetivo, consciência e intenção, esses modelos de responsabilidade não podem ser prontamente aplicados à IA em si porque não conseguem satisfazer a condição epistêmica necessária (Hildebrandt 2013; Himma 2009; Solum 1991; Gless *et al.*, 2016; Andrade *et al*., 2007 *apud* Yeung, 2018, p. 52).

No entanto, é possível que existam situações ondem os modelos baseados em intenção/culpabilidade possam ser aplicados aos desenvolvedores ou usuários humanos da IA. Isso pode ocorrer quando os indivíduos intencionalmente utilizam as tecnologias para propósitos maliciosos, como por exemplo, para cometer fraude ou apropriar-se indevidamente de propriedade. Nesses casos, claramente os requisitos para estabelecer responsabilidade sob um modelo baseado em intenção/culpabilidade estariam satisfeitos.

Nestas circunstâncias, uma violação prima facie dos direitos humanos surgiria e também seria possível gerar responsabilidade tanto no direito penal por crimes contra a pessoa (ou propriedade) quanto acionar obrigações civis de reparação e restauração (Hallevy, 2015).

3.6.2 Modelos baseados no risco e na negligência

Em alguns sistemas jurídicos, os modelos de responsabilidade baseados no risco e na negligência, em caso de danos tangíveis, formam a base de um dever geral de cuidado, buscando a prevenção de riscos que potencialmente poderiam

causar prejuízos. Esses modelos de responsabilidade são convencionalmente aplicados para determinar se os agentes estão sujeitos a obrigações legais de reparação para aqueles que sofreram danos como resultado da falha do agente em cumprir esse dever geral de cuidado (Yeung, 2018, p. 52).

Uma condição de controle semelhante à que se aplica aos modelos baseados em intenção/culpabilidade de responsabilidade também se aplica aos modelos baseados em risco/negligência. No entanto, a condição epistêmica aplicável aos modelos baseados em risco/negligência é, em alguns sistemas jurídicos, consideravelmente menos exigente do que aquelas aplicáveis aos modelos baseados em intenção/culpabilidade (Yeung, 2018, p. 52).

Conforme explica John Oberdiek (2017, p. 57), fatos podem ter importância moral: eles possuem uma força normativa que incide sobre a permissibilidade da ação prospectiva, mas apenas depois de serem razoavelmente descobertos.

Ao decidir sobre um curso de ação, Oberdiek (2017, p. 57) destaca que pode ser moralmente esperado que uma pessoa comum tome cuidado epistêmico razoável. Segundo o autor, ela não pode ser obrigada a conhecer todos os fatos, mas também não pode fechar os olhos e se basear em sua compreensão subjetiva, deixando, assim, de tomar a razoável cautela para descobrir ou conhecer fatos relevantes.

Portanto, se a responsabilidade com base em um modelo de risco/negligência pode ser atribuída aos desenvolvedores humanos de agentes computacionais e sistemas dependerá se esse dano era uma consequência razoavelmente previsível das ações e decisões dos sistemas computacionais. Nesse sentido, o dever de cuidado surge quando, falando de maneira muito ampla, existe um risco razoavelmente previsível de que uma ação possa prejudicar uma pessoa e, ainda assim, o agente decida prosseguir com a ação (Yeung, 2018, p. 53).

A previsibilidade, portanto, opera tanto para definir os tipos de riscos pelos quais uma pessoa pode ser legalmente responsável quanto para limitar os danos pelos quais ela pode ser responsabilizada (Yeung, 2018, p. 53).

A previsibilidade razoável também desempenha um papel na determinação de como se espera que uma pessoa aja. O dever de cuidado é cumprido se uma pessoa age como uma pessoa comum exercendo cuidado razoável para evitar riscos previsíveis (Oberdiek, 2017, p. 40). Portanto, a previsibilidade razoável

funciona como o critério para avaliar se as atividades envolvem risco e se podem resultar em danos tangíveis a outros, gerando um dever legal de cuidado.

Como observa Oberdiek (2017, p. 48), no padrão da *common law*, por exemplo, no caso de atividades de risco, é importante que se possa responsabilizar uns aos outros por suas respectivas definições de risco. Em outras palavras, as pessoas devem ser capazes de justificar os seus riscos de uma maneira que suporte o escrutínio moral.

No entanto, para identificar se é razoavelmente previsível que qualquer ação arriscada possa resultar em dano, nos deparamos com o chamado "problema da classificação da referência". Como explica Oberdiek, “o problema da classificação de referência é essencialmente um problema de redescrição - qualquer risco específico pode ser infinitamente redesenhado; não há uma classificação única que classifique as crenças de acordo com os seus objetos" (Oberdiek, 2017, p. 40).

Por exemplo, considere a lesão fatal causada por um veículo Uber que colidiu com uma mulher que empurrava uma bicicleta com sacolas de compras penduradas no guidão em 2018. O veículo estava operando no modo de direção autônoma por 19 minutos antes de confundir a mulher com um carro, e reconhecendo seu erro e devolvendo o controle ao motorista humano do veículo segundos antes da colisão, que o motorista humano não conseguiu evitar (Smith, 2018).

Parece improvável que os desenvolvedores do carro pudessem prever razoavelmente que o sistema de IA do veículo acreditaria erroneamente que uma mulher que empurra uma bicicleta com sacolas de compras penduradas no guidão era outro veículo. Por outro lado, parece estar dentro dos limites da previsão razoável que as tecnologias de detecção do carro falhariam em classificar corretamente objetos de formatos incomuns encontrados durante condições normais de direção e que erros desse tipo poderiam levar a colisões fatais (Yeung, 2018, p. 54).

Ao mesmo tempo, identificar se eventos específicos associados à operação de um objeto tecnológico específico são "razoavelmente previsíveis" é inevitável. Nas fases iniciais, quando uma nova tecnologia está sendo implementada, as expectativas sobre seus comportamentos (e consequências) serão relativamente instáveis e desconhecidas, no entanto, com o tempo e à medida que se acostuma com os padrões de comportamento e ação, esses comportamentos e ações podem

se tornar mais familiares aos desenvolvedores e, portanto, mais propensos a serem considerados razoavelmente previsíveis (Yeung, 2018, p. 55).

Portanto, os desenvolvedores dessas tecnologias devem ser responsabilizados por negligência ao deixar de tomar medidas que teriam evitado o dano decorrentes da infração (Liu; Zawieska, 2020). Mesmo assim, isso levanta a questão sobre as expectativas da indústria de tecnologia ao tomar a decisão de lançar uma tecnologia emergente em contextos do mundo real (Yeung, 2018, p. 55).

Outras questões surgem sobre o padrão mínimo de cuidado ao qual os desenvolvedores de sistemas de IA devem se atentar. Considerando novamente a colisão fatal do veículo Uber que classificou erroneamente um pedestre que empurrava uma bicicleta como um veículo se aproximando, resta um questionamento importante: seria apropriado aplicar o mesmo modelo de responsabilidade e o mesmo padrão de cuidado que se aplica a um motorista comum operando um carro dirigido por humanos tradicionais aos danos não intencionados resultantes das ações de um carro autônomo? Em outras palavras, há escolhas políticas importantes a serem feitas e não é, de forma alguma, claro e evidente que o padrão do motorista humano comum seja a comparação mais adequada a ser feita (Yeung; Howes; Pogrebna, 2020).

3.6.3 Responsabilidade estrita

Como observado anteriormente, o modelo de responsabilidade legal aplicável a violações de direitos (incluindo violações de direitos humanos e liberdades fundamentais) é o da responsabilidade estrita. Neste modelo, a responsabilidade é atribuída ao agente sem necessidade de prova de culpa, de modo que a responsabilidade legal por violações de direitos se aplica àqueles que as causam, independentemente de o agente responsável ter se envolvido em conduta que violou um padrão de conduta legalmente especificado, e independentemente de a conduta ter sido intencional (Cane, 2002, p. 82).

Das quatro variedades de responsabilidade estrita identificadas por Cane, três são de relevância direta para este estudo, quais sejam: responsabilidade baseada em direitos; baseada em resultados e baseada em atividade (Yeung, 2018, p. 56).

(a) Responsabilidade estrita baseada em direitos: surge quando direitos legais são violados, de modo que qualquer violação da esfera de proteção delimitada pelo direito aciona a responsabilidade (Yeung, 2018, p. 56).

(b) Responsabilidade estrita baseada em resultados: esta forma de responsabilidade repousa quando resultados adversos são causados independentemente de culpa. Em relação às tecnologias digitais avançadas, surgem questões sobre o que constitui um defeito relevante (Yeung, 2018, p. 57).

Considerando novamente a colisão fatal do veículo da Uber, que inicialmente classificou erroneamente um pedestre conduzindo uma bicicleta como outro veículo, devolvendo o controle ao motorista humano assim que reconheceu o erro, mas, no entanto, tarde demais para o motorista humano evitar a colisão. Pode-se argumentar que, nessas circunstâncias, o veículo não era defeituoso, na medida em que funcionava exatamente da maneira que seus desenvolvedores pretendiam. Por outro lado, se defeituoso for interpretado como adequado para o propósito, então a falha do veículo em classificar corretamente o pedestre e tomar medidas evasivas para evitar a colisão fatal poderia ser prontamente caracterizada como defeituosa (Yeung, 2018, p. 57).

Uma abordagem semelhante é frequentemente aplicada quando o risco de dano está ligado à imprevisibilidade do comportamento de grupos específicos de risco, como animais. Nesses casos, a responsabilidade é atribuída às pessoas consideradas responsáveis por supervisionar o animal, pois geralmente são consideradas as mais aptas a adotar medidas para prevenir ou reduzir o risco de dano (Yeung, 2018, p. 57).

(c) Responsabilidade estrita baseada em atividade surge em conexão com uma atividade específica, como leis que proíbem a posse de armas, facas, substâncias ilícitas e assim por diante. Em algumas jurisdições, pode-se adotar uma abordagem de responsabilidade estrita para aqueles que realizam atividades perigosas (por exemplo, o operador de uma usina nuclear ou de uma aeronave) são ultimamente responsáveis pela atividade perigosa (por exemplo, o proprietário de um veículo).

> Nesses casos, a justificativa subjacente é que essa pessoa criou um risco e, ao mesmo tempo, obtém um benefício econômico dessa atividade (Yeung, 2018, p. 57).

Essas várias formas de responsabilidade estrita distribuem os riscos associados a atividades potencialmente prejudiciais entre agentes e vítimas, dando considerável peso aos interesses das vítimas na segurança da pessoa e da propriedade. Ao fazer isso, reconhecem que a responsabilidade não é apenas uma consequência da vontade de um agente, mas também está relacionada ao interesse que todos compartilham na segurança da pessoa e da propriedade, bem como na forma como os recursos e riscos são distribuídos na sociedade, delineando assim os limites do que são as responsabilidades dos indivíduos (Cane, 2002, p. 109).

3.6.4 Seguro mandatório

Em vez de concentrar-se em atribuir responsabilidade a candidatos em potencial que possam ser entendidos como contribuintes para os danos e erros que podem surgir da operação de tecnologias digitais avançadas, uma sociedade pode decidir priorizar a necessidade de garantir que todos os prejudicados pela operação dessas tecnologias sejam financeiramente compensados. Isso pode ser alcançado instituindo um seguro obrigatório, que poderia ser estabelecido com base na ausência de culpa, criando um fundo de seguro ao qual todos os prejudicados pela operação dessas tecnologias poderiam recorrer (Yeung, 2018, p. 57).

De acordo com o Comitê, esse fundo pode ser financiado de várias maneiras, inclusive por meio de contribuições da indústria de tecnologia, contando com a administração de alguma instituição pública. Também seria possível simplesmente exigir que empresas envolvidas na cadeia de valor por meio da qual esses sistemas digitais avançados são projetados e implementados contratem um seguro obrigatório de responsabilidade.

Esse plano, apesar de não ser a única solução viável ou até mesmo o resultado perfeito e esperado em caso de dano, tem o benefício de permitir que aqueles prejudicados pela operação de tais tecnologias busquem compensação financeira em circunstâncias em que é difícil identificar precisamente quais empresas

devem ser consideradas responsáveis pelo dano, garantindo que mesmo que as empresas se tornem insolventes, a compensação será recebida.

Isso pode se tornar cada vez mais importante à medida que a sociedade passa a depender mais de sistemas inteligentes autônomos que continuam a operar muito tempo depois que seus desenvolvedores e proprietários humanos ou corporativos morreram ou deixaram de existir, de modo que as sociedades podem precisar desenvolver instituições de proteção de último recurso, como seguros coletivos, para garantir que as vítimas não sejam sistematicamente deixadas sem compensação (Yeung, 2018, p. 57).

3.6.5 Desafios da responsabilidade

A análise anterior avançou em grande parte com a suposição de que, ao buscar atribuir responsabilidade pelas consequências adversas das tecnologias digitais avançadas, relações de causa e efeito podem ser prontamente identificadas.

Na prática, no entanto, essas tecnologias constituem um componente essencial de sistemas sócio técnicos altamente complexos e sofisticados, gerando desafios agudos ao tentar identificar linhas de responsabilidade causal, moral e legal. Três desses desafios são brevemente explicados pelo Comitê e delineados a seguir: o problema das “muitas mãos”, “humanos na interação” e os efeitos imprevisíveis de dinâmicas complexas que podem surgir entre múltiplos sistemas algorítmicos interativos (Yeung, 2018, p. 58).

Exceto em relação a algumas formas de responsabilidade estrita, a atribuição de responsabilidade pelas ameaças, riscos, danos e violações de direitos (incluindo violações de direitos humanos) geralmente exige uma avaliação de se podem ser entendidos como causados pelo agente. No entanto, ao buscar atribuir responsabilidade causal a algum evento adverso ou efeito que poderia plausivelmente ser considerado uma consequência direta da operação de qualquer sistema sócio-técnico complexo, quer ele utilize ou não tecnologias de IA, depara-se imediatamente com o problema das “muitas mãos” (Yeung, 2018, p. 58).

Este problema surge ao adotar um modelo de responsabilidade baseado em intenção/culpabilidade. Primeiramente identificado no contexto da tecnologia da informação pela filósofa da tecnologia, Helen Nissenbaum, o problema das “muitas mãos” não é exclusivo de computadores, tecnologia digital, algoritmos ou

aprendizado de máquina. Pelo contrário, refere-se ao fato de que uma complexa variedade de indivíduos, organizações, componentes e processos estão envolvidos no desenvolvimento e implementação de sistemas complexos, tornando-se muito difícil identificar quem é o culpado quando esses sistemas apresentam mau funcionamento ou causam danos (Nissenbaum, 1996 *apud* Yeung, 2018, p. 58).

Isso ocorre porque esses conceitos são convencionalmente compreendidos em termos de concepções individualistas de responsabilidade. Em outras palavras, a responsabilidade causal é necessariamente distribuída quando se trata de sistemas tecnológicos complexos, diluindo a causalidade para meramente influência (Yeung, 2018, p. 58).

O problema das “muitas mãos” pode ser especialmente agudo ao tentar identificar o local da responsabilidade por danos ou erros resultantes do desenvolvimento e operação de sistemas de IA, uma vez que eles dependem de vários componentes críticos, a saber:

(a) Os modelos que são desenvolvidos para representar o espaço de características e o objetivo de otimização que o sistema se destina a alcançar;
(b) Algoritmos, baseados nesses modelos, que analisam os dados para produzir resultados que podem desencadear algum tipo de ação ou decisão;
(c) Os dados de entrada, que podem ou não incluir dados pessoais nos quais esses algoritmos são treinados;
(d) Os desenvolvedores humanos envolvidos no design desses sistemas, que devem tomar decisões repletas de valores sobre os modelos, algoritmos e dados que são usados para treinar os algoritmos nos quais o desempenho é testado. Isso inclui seres humanos que realizam a tarefa de rotular os dados usados para treinar os algoritmos;
(e) O sistema sócio técnico e o contexto mais amplo nos quais o sistema algorítmico está incorporado e no qual opera.

Mesmo supondo que seria possível identificar satisfatoriamente a alocação de responsabilidade moral por eventuais danos relacionados a cada um dos componentes acima, isso provavelmente não garantiria que fosse possível

facilmente aplicar a responsabilidade moral em situações onde não houve a intencionalidade, levando em consideração todo o contexto e complexidade do sistema. Esses desafios são agravados pelo fato de os produtos e serviços digitais estarem sujeitos a extensões de software, atualizações e correções após sua implementação (Yeung, 2018, p. 58).

Qualquer alteração no software do sistema pode afetar o comportamento de todo o sistema ou de componentes individuais, estendendo sua funcionalidade. Essas alterações podem modificar o perfil de risco operacional do sistema, incluindo sua capacidade de operar de maneiras que podem causar danos ou violar direitos humanos (Yeung, 2018, p. 58).

Ao lidar com esses desafios, pode ser útil ter em mente três considerações. Em primeiro lugar, questões relacionadas à alocação de responsabilidade legal por danos decorrentes de atividades envolvendo múltiplas partes não são novas, e muitos sistemas legais desenvolveram, portanto, um conjunto relativamente sofisticado de princípios e procedimentos para determinar a responsabilidade quando estão envolvidos múltiplos potenciais réus (Yeung, 2018, p. 58).

Como observou recentemente o Comitê, identificar a distribuição de responsabilidade para reparação entre vários atores envolvidos na cadeia de valor por meio da qual as tecnologias digitais emergentes operam pode não ser relevante para garantir que as vítimas obtenham compensação pelos danos sofridos, embora resolver tais questões seja provavelmente importante para fornecer segurança jurídica àqueles envolvidos na produção e implementação dessas tecnologias (Yeung, 2018, p. 58).

Em segundo lugar, e relacionado a isso, a capacidade da lei de elaborar respostas práticas, apesar da aparente intratabilidade do problema das muitas mãos, pode ser atribuída pelo menos em parte à ênfase maior que ela coloca nos interesses legítimos do paciente moral, em vez do foco quase exclusivo no agente moral, como ocorre nas teorias de escolha de responsabilidade moral (Yeung, 2018, p. 58).

Em terceiro lugar, é especialmente importante garantir que se tenha mecanismos eficazes e legítimos que operarão para prevenir e antecipar violações de direitos humanos, especialmente dado que muitas violações de direitos humanos associadas à operação de tecnologias digitais avançadas podem não resultar em danos tangíveis à saúde ou propriedade individual. A necessidade de uma

abordagem preventiva é especialmente importante dada a velocidade e escalabilidade com que essas tecnologias operam (Yeung, 2018, p. 58).

Os efeitos cumulativos e agregados das violações de direitos humanos causadas pela operação de sistemas de IA podem erodir seriamente as bases sociais necessárias para ordens morais e democráticas que são condições essenciais para que os direitos humanos existam, sugerindo que as abordagens existentes para a proteção dos direitos humanos podem precisar ser revitalizadas em uma era de rede e dados (Yeung, 2018, p. 58).

Não apenas muitos indivíduos, empresas e outras organizações estão envolvidos no desenvolvimento e implementação de tecnologias digitais avançadas, mas essas tecnologias muitas vezes são projetadas para operar de maneiras que envolvem a manutenção de envolvimento humano ativo. Isso aponta para sérios desafios associados à identificação da distribuição apropriada de autoridade e responsabilidade entre humanos e máquinas, dada a complexa interação entre eles (Yeung, 2018, p. 58).

Em particular, muitas tarefas anteriormente realizadas por humanos agora são executadas por máquinas, mas os humanos estão invariavelmente envolvidos em vários pontos ao longo da cadeia de desenvolvimento, teste, implementação e operação. Como observou a *Royal Academy of Engineering*: "Haverá sempre humanos na cadeia, mas não está claro, no caso de danos, qual humano na cadeia é responsável - o designer, fabricante, programador ou usuário" (Royal Academy Of Engineering, 2009, p. 2 *apud* Yeung, 2018, p. 58)

A interação entre humanos e máquinas dentro de sistemas sócio técnicos complexos e dinâmicos gera questões especialmente desafiadoras sobre o papel dos humanos na supervisão de operação da máquina. Uma preocupação recorrente tem sido a de que, para garantir que sistemas sócio técnicos cada vez mais complexos sempre operem a serviço da humanidade, esses sistemas devem sempre ser projetados para que possam ser desligados por um operador humano. No entanto, como observou novamente a Royal Academy of Engineering:

> Pode-se pensar que sempre há necessidade de intervenção humana, mas às vezes são necessários sistemas autônomos onde os humanos podem tomar decisões ruins como resultado do pânico - especialmente em situações estressantes - e, portanto, a intervenção humana seria problemática. Operadores humanos nem sempre estão certos e nem sempre têm as melhores intenções. Poderiam sistemas autônomos ser mais

> confiáveis do que operadores humanos em algumas situações? (Royal Academy Of Engineering, 2009, p. 4 *apud* Yeung, 2018, p. 58).

Por outro lado, mesmo que os humanos sejam mantidos no circuito com o objetivo de supervisionar sistemas computacionais, indivíduos colocados nessas posições podem estar compreensivelmente relutantes em intervir. Há mais de uma década, Johnson e Powers (2005, p. 106 *apud* Yeung, 2018, p. 60) comentaram:

> No caso do futuro controle automatizado de tráfego aéreo [...] haverá uma questão difícil sobre se e quando os controladores de tráfego aéreo humanos devem intervir no controle por computador de aeronaves [...] Aqueles humanos que anteriormente tinham a responsabilidade serão substituídos por cuidadores da tecnologia. Uma preocupação neste ambiente é que os humanos designados para interagir com esses sistemas 'automáticos' podem perceber a intervenção como moralmente arriscada. É melhor, eles podem raciocinar, permitir que o sistema de computador aja e que os humanos fiquem fora do caminho. Intervir no comportamento de sistemas automatizados de computador é colocar em dúvida a sabedoria dos designers do sistema e a 'experiência' do sistema em si. Ao mesmo tempo, uma pessoa que escolhe intervir no sistema traz sobre si o peso significativo da responsabilidade moral, e, portanto, os controladores humanos têm algum incentivo para deixar a automaticidade do sistema de computador lidere e toque o processo. Isso é uma fuga da responsabilidade por parte dos humanos, e mostra como a responsabilidade foi, de alguma forma, atribuída ao sistema de computador.

No entanto, à medida que se depende cada vez mais da ampla gama de serviços e sistemas automatizados, especialmente à medida que as tecnologias digitais se tornam cada vez mais poderosas e sofisticadas, a insistência contínua em colocar um humano no circuito para atuar em supervisionando, acaba por transformar os humanos em objetos morais absorvedores de culpa, mesmo que tenham apenas controle parcial do sistema, e sendo vulneráveis (Yeung, 2018, p. 61).

Como destaca o estudo de Elish e Twang sobre litígios envolvendo pilotos automáticos de aviação, as aeronaves modernas são controladas em grande parte por software, mas os pilotos nos cockpits continuam sendo legalmente responsáveis pela operação da aeronave. No entanto, nossas percepções culturais tendem a exibir um “viés de automação”, elevando a confiabilidade e infalibilidade da tecnologia automatizada enquanto culpam os humanos por erros (Elish, 2016 *apud* Yeung, 2018, p. 62).

De acordo com o Comitê, desafios ainda mais complexos surgem ao tentar identificar, antecipar e prevenir eventos adversos resultantes das interações entre

sistemas sócio técnicos complexos impulsionados por algoritmos, que podem ocorrer em uma velocidade e escala que simplesmente não eram possíveis em uma era pré-digital. O chamado 'flash crash' que ocorreu em 2010, durante o qual o mercado de ações entrou em queda livre por cinco minutos antes de se corrigir, sem motivo aparente, fornece uma ilustração vívida.

Agentes de IA, que têm a capacidade de aprender com seu ambiente e aprimorar iterativamente seu desempenho, podendo ser submetidos à verificação e teste matemáticos, trabalhando com diferentes algoritmos em um ecossistema complexo e dinâmico geram riscos de resultados imprevisíveis e potencialmente perigosos. Em outras palavras, essas interações geram riscos que o ser humano ainda pode não ser capaz de plenamente compreender.

O desafio de desenvolver soluções que nos permitam prever, modelar e agir de maneira confiável para evitar resultados indesejados e potencialmente catastróficos decorrentes da interação entre sistemas sócio técnicos dinâmicos e complexos cria uma nova e cada vez mais urgente fronteira para a pesquisa computacional. Os principais cientistas da computação Shadbolt e Hampson alertam para os perigos de "sistemas hipercomplexos e super-rápidos" gerando consideráveis novos riscos, e para os quais: "Nossa resposta precisa ser vigilante, inteligente e inventiva. Enquanto formos, permaneceremos no controle das máquinas e nos beneficiaremos muito delas. Precisamos desenvolver estruturas de políticas para isso. Além dos perigos, surge um mundo de oportunidades." (Shadbolt, Hampson, 2018 *apud* Yeung, 2018, p. 62).

3.7 Responsabilidade do Estado em garantir a proteção efetiva dos direitos humanos

Uma das preocupações mais significativas sobre o surgimento de sistemas algorítmicos tem sido o aumento do poder das grandes empresas de tecnologia, incluindo preocupações sobre a grande assimetria de poder entre essas empresas e os indivíduos que estão sujeitos a elas. No entanto, apesar de o poder de implantar sistemas algorítmicos estar nas mãos dessas grandes empresas, a obrigação de proteger os direitos humanos no âmbito internacional recai principalmente sobre os Estados, uma vez que a proteção dos direitos humanos é principalmente destinada a

operar verticalmente, para proteger os indivíduos contra interferências injustificadas (Yeung, 2018, p. 62).

Vale ressaltar ainda que de acordo com o Comitê, é amplamente estabelecido na jurisprudência da CEDH (Convenção Europeia dos Direitos Humanos) que os direitos protegidos pela Convenção fundamentam obrigações substantivas positivas que exigem que os Estados-membros ajam para garantir que aqueles dentro de sua jurisdição tenham os seus direitos protegidos (Rainey; Wicks; Ovey, 2014, p. 102 *apud* Yeung, 2018, p. 62).

Portanto, os Estados são obrigados pela CEDH a introduzir legislação nacional e outras políticas necessárias para garantir que os preceitos da CEDH sejam devidamente respeitados, incluindo a proteção contra interferências de outros (incluindo empresas de tecnologia), que podem, portanto, estar sujeitos a deveres legais vinculativos de respeitar os direitos humanos (Yeung, 2018, p. 63).

Essas obrigações são exigíveis e estão fundamentadas na proteção dos direitos humanos pela Convenção, incluindo o direito a um recurso efetivo, que ofereça bases sólidas para impor mecanismos legalmente exigíveis e eficazes, para garantir responsabilidade por violações dos direitos humanos. Estes termos estão bem além daquilo que a retórica contemporânea da ética da IA na forma de autorregulação voluntária pela indústria de tecnologia pode realisticamente oferecer (Yeung, 2018, p. 63).

A discussão de vários modelos para alocar responsabilidade histórica, delineada em capítulos anteriores, baseia-se amplamente em abordagens legais e julgados de tribunais, utilizando sua interpretação e aplicação do direito consuetudinário para determinar a responsabilidade legal por danos. Uma desvantagem significativa associada à dependência de remédios judiciais para remediar essas preocupações é que eles são mais adequados para remediar danos substanciais sofridos por poucos, em oposição a danos menos significativos sofridos por muitos (Yeung, 2018, p. 63).

As dificuldades em buscar reparação por meio dos tribunais são ampliadas no espaço da IA pelo desafio de detectar o dano e determinar e provar a causalidade, sem mencionar os sérios obstáculos práticos e desincentivos enfrentados pelos indivíduos ao invocar o processo judicial (Mantelero, 2018, p. 55 *apud* Yeung, 2018, p. 63).

Ao mesmo tempo, a capacidade dos sistemas de IA em um ambiente globalmente interconectado de gerar problemas de ação coletiva já foi destacada, frisando a necessidade e a importância de uma fiscalização nacional devidamente equipada com autoridades e poderes de fiscalização adequados (Yeung, 2018, p. 64).

Ao mesmo tempo, é importante reconhecer que, além dos mecanismos legais convencionais de reparação por meio dos tribunais, existem muitos outros mecanismos institucionais de governança que poderiam ajudar a garantir o desenvolvimento e a implementação responsáveis e conformes aos direitos humanos de tecnologias digitais avançadas. Por isso, a seguir será trazido um breve resumo de outros possíveis mecanismos de governança institucional (além das iniciativas autorreguladas voluntárias atualmente emergentes) que podem servir para aprimorar tanto a responsabilidade prospectiva quanto retrospectiva pelas ameaças, riscos, danos e condutas prejudiciais decorrentes da operação de tecnologias digitais avançadas (Yeung, 2018, p. 64).

3.8 Mecanismos não judiciais para fazer cumprir a responsabilidade pelas tecnologias digitais avançadas

Embora os mecanismos de governança regulatória possam ser classificados de várias maneiras, três características são dignas de destaque para os fins deste estudo, de acordo com o Comitê. Em primeiro lugar, podemos distinguir entre mecanismos que operam de forma *ex ante*, fornecendo supervisão e avaliação de um objeto, processo ou sistema antes de ser implementado em ambientes do mundo real e, portanto, preocupados principalmente em garantir a responsabilidade prospectiva. Por outro lado, existem mecanismos *ex post* que operam durante ou após a implementação e, portanto, estão preocupados principalmente em garantir a responsabilidade histórica (Yeung, 2018, p. 64).

Como enfatizado anteriormente, ambas as dimensões de responsabilidade devem ser consideradas para garantir o desenvolvimento e a implementação responsáveis de sistemas de IA. No entanto, porque este estudo está preocupado principalmente com as implicações dos direitos humanos dessas tecnologias, a necessidade de mecanismos eficazes e legítimos que evitem violações dos direitos humanos é de considerável importância, especialmente dada a velocidade e escala

com que os sistemas de IA podem operar agora, combinados com uma cultura de “mover-se rapidamente e quebrar paradigmas” que caracteriza a estratégia operacional das principais empresas de tecnologia (Yeung, 2018, p. 64).

Esta estratégia consiste em avançar com inovação tecnológica rápida sem atender cuidadosamente aos riscos potenciais antecipadamente, preferindo lidar com qualquer resultado adverso após o lançamento da tecnologia, momento em que pode não ser praticamente possível desfazer ou reverter as inovações tecnológicas que já foram lançadas no mercado, do que perder a oportunidade de desenvolver algo disruptivo (Taplin, 2018; Vaidhyanathan, 2011 *apud* Yeung, 2018, p. 63).

Em segundo lugar, é importante atender à legalidade das instituições e mecanismos de governança regulatória para identificar se, e em que medida, são considerados mecanismos opcionais que a indústria de tecnologia tem a liberdade de adotar seletivamente ou ignorar completamente, ou se são legalmente mandatórios e para os quais sanções substanciais estão associadas à não conformidade (Nemitz, 2018 *apud* Yeung, 2018, p. 64).

Em terceiro lugar, embora os mecanismos de governança regulatória tenham tomado convencionalmente a forma de instituições sociais, no contexto atual, o papel de mecanismos técnicos de proteção, que dependem de uma modalidade de controle às vezes chamada de “regulação by design”, pode ser igualmente (se não mais) importante (Yeung, 2015 *apud* Yeung, 2018, p. 64).

3.8.1 Técnicas e instrumentos regulatórios

Formas mais convencionais, sociais e organizacionais de instrumentos de governança regulatória também surgiram em resposta ao reconhecimento de que as tecnologias de IA podem prejudicar valores importantes, objetivando garantir que esses sistemas tecnológicos operem de maneiras que respeitem os direitos humanos. Dois deles merecem maior destaque: (1) a avaliação de impacto de direitos humanos; e (2) as técnicas de auditoria algorítmica (Yeung, 2018, p. 65-66).

No contexto da Avaliação de Impacto Algorítmico em relação aos Direitos Humanos, vários estudiosos e organizações têm apresentado diversas abordagens para avaliar o impacto de algoritmos. Essas abordagens, na prática, se assemelham a modelos de avaliação de riscos destinados a serem utilizados por aqueles que desejam adquirir ou implementar sistemas algorítmicos. O objetivo é identificar as

implicações éticas, sociais e de direitos humanos de seus sistemas propostos e tomar medidas para mitigar essas preocupações no design e na operação dos sistemas algorítmicos antes de sua implementação. Além dos modelos gerais de avaliação de impacto, também foram propostos modelos específicos para diferentes áreas de atuação (Mantelero, 2018 *apud* Yeung, 2018, p. 65).

Esses modelos de avaliação de risco variam amplamente em termos de: (a) critérios de avaliação; (b) parte responsável pela avaliação; (c) adoção obrigatória ou voluntária; e (d) escala de avaliação (Yeung, 2018, p. 65-66).

O uso de critérios de avaliação desempenha um papel fundamental na análise do impacto de sistemas algorítmicos. Enquanto a legislação de proteção de dados da União Europeia requer a utilização de Avaliações de Impacto de Proteção de Dados (DPIAs) em determinadas situações, outros modelos, como a Avaliação de Impacto de Direitos Humanos, buscam avaliar o impacto de sistemas propostos nos direitos humanos de forma mais abrangente. Além disso, a responsabilidade pela avaliação pode variar, sendo realizada pelo controlador de dados, por uma terceira parte externa ou por um órgão de credenciamento, dependendo do modelo adotado (Yeung, 2018, p. 65-66).

A adoção dessas técnicas de avaliação pode ser voluntária ou obrigatória, com alguns modelos permitindo que o controlador de dados escolha realizar a avaliação e tomar medidas com base nela, enquanto outros defendem a exigência legal. Ademais, a escala de avaliação varia, com a Avaliação de Impacto de Direitos Humanos abrangendo uma ampla gama de operações comerciais para avaliar a conformidade com os padrões de direitos humanos, enquanto outras abordagens são mais específicas, concentrando-se em atividades individuais de processamento de dados (Yeung, 2018, p. 65-66).

No entanto, é importante destacar que, embora essas técnicas de avaliação sejam valiosas para identificar riscos de interferência nos direitos humanos, é necessário desenvolver abordagens metodológicas rigorosas e consistentes para garantir que as organizações as adotem de maneira eficaz. Isso requer um compromisso genuíno em identificar e mitigar riscos de direitos humanos, em vez de considerar a conformidade apenas como um procedimento burocrático sem uma preocupação real com o respeito aos direitos humanos (Yeung, 2018, p. 65-66).

Como explicado no Capítulo 1, o Projeto de Lei 2338/23 que está em tramitação no Congresso Nacional, apesar de adotar a avaliação de impacto, não

aborda de maneira a demonstrar um interesse genuíno na identificação dos riscos de direitos humanos, o que pode se tornar perigoso (Yeung, 2018, p. 66).

Já a Auditoria Algorítmica representa uma abordagem distinta em relação às avaliações de impacto, ocorrendo após a implementação do sistema. Essas técnicas são projetadas para testar e avaliar sistemas algorítmicos já em funcionamento. Este campo emergente de pesquisa técnica se fundamenta em ferramentas e técnicas em desenvolvimento, destinadas a detectar, investigar e diagnosticar quaisquer efeitos indesejados de sistemas algorítmicos (Yeung, 2018, p. 66-67).

O Comitê propõe a formalização e institucionalização de técnicas desse tipo por meio de um quadro regulatório legalmente mandatório. Isso se aplicaria aos sistemas algorítmicos, especialmente aqueles considerados de alto risco devido à gravidade e escala de possíveis consequências em caso de falha ou efeitos adversos não intencionados. Esses sistemas seriam submetidos a revisões e supervisões regulares realizadas por uma autoridade externa composta por especialistas técnicos qualificados. Por exemplo, Cukier e Mayer-Schonenberg (Mayer-Schonberger; Cukier, 2013, p. 180 *apud* Yeung, 2018, p. 67) sugerem a necessidade de uma nova categoria profissional, os "algoritmistas", que desempenhariam um papel semelhante ao de profissionais de áreas como direito, medicina, contabilidade e engenharia. Os algoritmistas poderiam ser independentes e externos para monitorar algoritmos de fora ou internos, contratados pelas organizações para monitorar os algoritmos desenvolvidos e implantados internamente, que seriam, então, sujeitos a revisões externas (Yeung, 2018, p. 67).

Sendo assim, mesmo que diversas sejam as maneiras de tratar a responsabilidade civil, resta claro que a criação de uma legislação específica sobre esse tema é extremamente importante. Entretanto, deve-se ter em mente que essa legislação primeiramente tem a obrigação de tecnicamente estar alinhada com o que os estudiosos do assunto entendem como inteligência artificial e como classificação de sistemas de alto risco.

4 CONSIDERAÇÕES FINAIS

A presente dissertação teve como objetivo compreender o que seria inteligência artificial, demonstrando o que, no momento, vem sendo mais discutido e pesquisado quando se trata do uso dessa tecnologia. Além disso, abordou a necessidade de uma regulação consistente e factível de ser cumprida, descrevendo o Projeto de Lei nº 2338/2023 e alguns parâmetros internacionais para essa regulação.

Ademais, foi possível verificar que, na prática jurídica também é possível fazer uso de novas ferramentas. Essas inovações, apesar de muito questionadas e do receio que muitos profissionais possuem de serem substituídos por máquinas, são inevitáveis e a melhor maneira de agir é aliando-se a elas, entendendo e estudando as diversas formas de tê-la como parceira.

Para os advogados, poder ter tempo hábil para dedicar-se à atividades intelectualmente mais complexas, tendo acesso a uma vasta quantidade de dados e análises é essencial e, pode elevar o nível dos serviços prestados e do gerenciamento dos riscos legais envolvidos em transações complexas. Essa seria mais uma demonstração de que os bacharéis em direito não são somente pessoas que criam burocracias e usam uma linguagem complexa para que ninguém os entenda.

A verdade é que quanto maior o engajamento da comunidade jurídica com a inteligência artificial, melhor será o resultado, tanto para a sociedade, quanto para os profissionais de direito. Poder contar com o apoio e com o conhecimento técnico nos debates sobre os riscos do uso da inteligência artificial pode alavancar os níveis de segurança das ferramentas, ao passo que garante maior assertividade, velocidade e confiabilidade às decisões tomadas pelos humanos.

Obviamente, a inteligência artificial representa um marco no cenário tecnológico contemporâneo, oferecendo inúmeros pontos positivos que permeiam diversos setores da sociedade. Um dos maiores benefícios é a otimização de processos. A IA tem a capacidade de analisar grandes volumes de dados rapidamente, identificando padrões e tendências que escapariam à percepção humana. Essa capacidade analítica aprimorada não apenas agiliza operações, mas também permite uma tomada de decisão mais informada e precisa.

Outro ponto positivo destacado pela utilização da IA é a capacidade de processar e analisar dados em larga escala. Em setores como a medicina, a IA é fundamental para a interpretação de imagens médicas, diagnósticos precoces e personalização de tratamentos com base em dados genéticos. Essa abordagem personalizada não apenas melhora os resultados clínicos, mas também representa um avanço significativo na prestação de cuidados de saúde.

Contudo, como exposto no decorrer da dissertação, é crucial abordar questões éticas e de segurança associadas à IA. Questões de privacidade dos dados tornam-se proeminentes à medida que algoritmos avançados analisam grandes conjuntos de informações pessoais. O acesso indiscriminado e o uso indevido desses dados podem comprometer a privacidade individual, levantando sérias questões éticas sobre quem detém e controla essas informações.

Outro aspecto crítico é o viés algorítmico. Algoritmos de IA aprendem com conjuntos de dados históricos, e se esses dados refletirem preconceitos existentes, a IA pode perpetuar e até ampliar essas discrepâncias. Isso levanta preocupações sobre a equidade e justiça na aplicação de sistemas automatizados.

A opacidade dos sistemas de IA é uma questão adicional. Modelos complexos e algoritmos de aprendizado profundo muitas vezes operam como caixas-pretas, dificultando a compreensão total de suas decisões. Isso levanta a questão da responsabilidade e prestação de contas, pois a falta de transparência pode tornar difícil identificar e corrigir falhas ou comportamentos indesejados.

A segurança cibernética é uma preocupação constante quando se trata de implementações de IA. Sistemas de IA podem se tornar alvos atrativos para ataques, levando a consequências sérias, como manipulação de dados, sabotagem ou espionagem. A dependência crescente da IA em setores críticos, como saúde e infraestrutura, aumenta a importância de salvaguardar esses sistemas contra ameaças cibernéticas.

Em última análise, por mais que a IA ofereça inúmeras vantagens e que deva ser utilizada em favor da sociedade, é importante aprofundar no estudo desses pontos negativos de maneira proativa. Desenvolver regulamentações éticas, promover a transparência dos algoritmos, investir em segurança cibernética robusta e adotar medidas para mitigar o impacto social da automação são passos essenciais para orientar o desenvolvimento e implementação responsável da inteligência artificial.

REFERÊNCIAS

ADELSON, Rachel. Marine mammals master math. **American Psychological Associaton**, Science watch, v. 36, n. 8, set., 2005. Disponível em: https://www.apa.org/monitor/sep05/marine. Acesso em: 02 jan. 2023.

BRASIL. Senado Federal. **Projeto de Lei nº 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial, 23 out. 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 23 out. 2023.

CANE, Peter. **Responsibility in law and morality**. Bloomsbury Publishing, 2002.

DANAHER, John. Robots, law and the retribution gap. **Ethics and Information Technology**, v. 18, n. 4, p. 299-309, 2016.

DEWAR, Robert B.K. COTS Software in critical systems: the case for Freely Licensed Open Source Software. **Military Embedded Systems**, 09 dez. 2010. Disponível em: https://militaryembedded.com/avionics/software/cots-open-source-software. Acesso em: 11 dez. 2022.

ENGELMANN, Wilson; WERNER, Deivid Augusto. Inteligência artificial e Direito. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). **Inteligência artificial e Direito**: ética, regulação e responsabilidade. 2. ed. rev., atual. e ampl. São Paulo: Thomson Reuters; Revista dos Tribunais, 2020, p. 145-174.

EUDES, Yves. The journalists who never sleep. **The Guardian**, 12 set. 2014. Disponível em: https://www.theguardian.com/technology/2014/sep/12/artificial-intelligence-data-journalism-media. Acesso em: 31 out. 2023.

FREZ, Célia Iaroz. Intertexto – Bertold Brecht (1898-1956). **Pet Letras**, Unicentro, 29 mar. 2017. Disponível em: https://www2.unicentro.br/pet-letras/2017/03/29/intertexto-bertold-brecht-1898-1956/. Acesso em: 31 out. 2023.

GRAEF, Aileen. Elon Musk: We are ‘summoning a demon’ with artificial intelligence. **UPI**, Business News, 27 out. 2014. Disponível em: https://www.upi.com/Business_News/2014/10/27/ElonMusk-We-are-summoning-a-demon-with-artificial-intelligence/4191414407652. Acesso em: 31 out. 2023.

GREENE, Daniel; HOFFMANN, Anna Lauren; STARK, Luke. Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. **Hawaii International Conference On System Sciences (Hicss-52)**, 08 – 11 jan., Hawaii, 2019.

GUIMARAENS, Alphonsus de. **Pastoral aos crentes do amor e da morte**. Poesia completa. Rio de Janeiro: Nova Aguilar, 1977. p. 313-314. Disponível em: https://digital.bbm.usp.br/handle/bbm/7695. Acesso em: 31 out. 2023.

GUTIERREZ, Andrei. É Possível Confiar em Um Sistema de Inteligência Artificial? Práticas em Torno da Melhoria da Sua Confiança, Segurança e Evidências de

Accountability. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. (coord.) **Inteligência Artificial e Direito**: Ética, Regulação e Responsabilidade. São Paulo: Revista dos Tribunais, 2019.

HALLEVY, Gabriel. **Liability for crimes involving artificial intelligence systems**. New York, USA: Springer International Publishing, 2015.

HART, Herbert Lionel Adolphus. Punishment and responsibility: **Essays in the philosophy of law**. Oxford University Press, 2008.

HOGEMANN, Edna Raquel. O futuro do Direito e do ensino jurídico diante das novas tecnologias. **Revista Interdisciplinar de Direito**, [S.l.], v. 16, n. 1, p. 105-115, jun. 2018. ISSN 2447-4290. Disponível em: http://revistas.faa.edu.br/index.php/FDV/article/view/487. Acesso em: 16 fev. 2020.

KAPLAN, A.; HAENLEIN, M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. **Business Horizons**, v. 62, n. 1, jan./fev., 2018. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0007681318301393. Acesso em: 03 fev. 2023.

LIU, Hin-Yan; ZAWIESKA, Karolina. From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence. **Ethics and Information Technology**, v. 22, p. 321-333, 2020.

MATTHIAS, Andreas. The responsibility gap: Ascribing responsibility for the actions of learning automata. **Ethics and information technology**, v. 6, p. 175-183, 2004.

MCCARTHY, John**. What is Artificial Intelligence?**. Standford University, Computer Science Department, 12 nov. 2007, 15 p. Disponível em: http://www-formal.stanford.edu/jmc/whatisai.pdf. Acesso em: 31 out. 2023.

MCGINNIS, John O. Accelerating AI. **Public Law and Legal Theory Series**, Northwestern University School of Law, n. 10-12, 25 abr., 2010, 26 p. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1593851. Acesso em: 23 jun. 2023.

METZINGER, T. Ethics Washing Made in Europe. **Der Tagesspiegel**, v. 8, 2019. Disponível em: https://www. tagesspiegel. de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html. Acesso em: 21 jan. 2023.

NEMITZ, Paul. Constitutional democracy and technology in the age of artificial intelligence. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 376, n. 2133, p. 20180089, 2018.

OBERDIEK, John. **Imposing risk**: A normative framework. Oxford University Press, 2017.

OLIVER, Dawn. Law, politics and public accountability. The search for a new equilibrium. **Public Law**, p. 238-238, 1994.

RUSSEL, Stuart; NORVIG, Peter. **Artificial Intelligence**: a modern approach. 3. ed. Nova Jersey: Prentice Hall, 2010. Disponível em: https://web.cs.ucla.edu/~srinath/static/pdfs/AIMA.pdf. Acesso em: 13 mar. 2023.

SCOTT, Mark; ISAAC, Mike. Facebook restores iconic Vietnam War photo it censored for nudity. **The New York Times**, v. 9, 2016. Disponível em: https://www.nytimes.com/2016/09/10/technology/facebook-vietnam-war-photonudity.html. Acesso em: 07 out. 2022.

SCHERER, Matthew. Regulating Artificial Intelligence systems: risks, challenges, competencies and strategies. Harvard Journal of Law and Technology, 2016.

SMITH, Andrew. Franken-algorithms: the deadly consequences of unpredictable code. **The Guardian**, v. 30, 2018.

STEIBEL, Fabro; VICENTE Victor Freitas; DE JESUS, Diego Santos Vieira. Possibilidades e Pontenciais da utilização da Inteligência Artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. (coord.) **Inteligência Artificial e Direito**: Ética, Regulação e Responsabilidade. São Paulo: Revista dos Tribunais, 2019.

TEPEDINO, G.; DA GUIA SILVA, R. Desafios da inteligência artificial em matéria de responsabilidade civil. **Revista Brasileira de Direito Civil**, [S. l.], v. 21, n. 03, p. 61-86, 2019. Disponível em: https://rbdcivil.emnuvens.com.br/rbdc/article/view/465. Acesso em: 20 fev. 2020.

WAGNER, Ben. Ethics as an escape from regulation. **From "ethics-washing" to ethics-shopping?**, p. 84-88, 2018.

WALLACE, R. Jay. **Responsibility and the moral sentiments**. Harvard University Press, 1994.

WOODY, Carol; ELLISON, Robert J. Supply-Chain Risk Management: Incorporating Security into Software Development. **Software Engineering Institute**, Carnegie Mellon University, white paper, 02 jul. 2013. Disponível em: https://insights.sei.cmu.edu/library/supply-chain-risk-management-incorporating-security-into-software-development/. Acesso em: 11 dez. 2022.

YEUNG, Karen. **A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework**. MSI-AUT, Conselho Europeu, 9 nov. 2018, 94 p. Disponível em: https://ssrn.com/abstract=3286027. Acesso em: 10 nov. 2023.

YEUNG, Karen; HOWES, Andrew; POGREBNA, Ganna. AI Governance by Human Rights–Centered Design, Deliberation, and Oversight. **The Oxford handbook of ethics of AI**, p. 77-106, 2020.