UNIVERSIDADE DE BRASÍLIA — UnB
FACULDADE DE DIREITO
CURSO DE GRADUAÇÃO EM DIREITO

FÁBIO DE BARROS CORREIA GOMES FILHO

INTELIGÊNCIA ARTIFICIAL

CONCEITO, RISCOS E REGULAÇÃO

Brasília, DF
2023

FÁBIO DE BARROS CORREIA GOMES FILHO

INTELIGÊNCIA ARTIFICIAL

CONCEITO, RISCOS E REGULAÇÃO

Monografia apresentada como requisito parcial para a obtenção do grau de Bacharel em Direito pela Faculdade de Direito da Universidade de Brasília.

Brasília, DF
2023

FÁBIO DE BARROS CORREIA GOMES FILHO

INTELIGÊNCIA ARTIFICIAL

CONCEITO, RISCOS E REGULAÇÃO

Monografia aprovada como requisito parcial para a obtenção do grau de Bacharel em Direito pela Faculdade de Direito da Universidade de Brasília, pela banca examinadora composta por:

_____

Professor Alexandre Araújo Costa
(Orientador — Presidente)

_____

Professor Marcus Vinicius Kiyoshi Onodera
(Membro da Banca Examinadora)

_____

Professora Fernanda de Carvalho Lage
(Membro da Banca Examinadora)

de Barros Correia Gomes Filho, Fábio
dG633i Inteligência Artificial: Conceito, Riscos e Regulação / Fábio de Barros Correia Gomes Filho; orientador Alexandre Araújo Costa. -- Brasília, 2023.
63 p.

Monografia (Graduação - Direito) -- Universidade de Brasília, 2023.

1. Inteligência artificial. 2. Direito da Inteligência Artificial. 3. Riscos e regulação da inteligência artificial. 4. AI Act do Parlamento Europeu. 5. Projeto de Lei 2.338/2023. I. Araújo Costa, Alexandre, orient. II. Título.

AGRADECIMENTOS

A Deus; à minha esposa Alexandra; a meus pais, Fábio e Anna; ao orientador deste projeto, professor Alexandre Araújo Costa; ao professor Marcus Vinicius Kiyoshi Onodera; à professora Fernanda de Carvalho Lage; a todos os professores com quem tive a oportunidade de conviver e aprender na Faculdade de Direito da Universidade de Brasília.

"No que toca ao intelecto, é provável que se trate de algo mais divino e impassível."

(Aristóteles)

RESUMO

Este trabalho discorre a respeito de impactos e respectivos riscos decorrentes do uso de tecnologias baseadas em inteligência artificial, para então adentrar questões regulatórias que dependem da apreciação desses riscos. Para isso, em primeiro lugar, dispersam-se equívocos comuns a respeito da temática, de forma a estabelecer definições precisas sobre inteligência artificial. A partir daí, passa-se à apresentação de uma tipologia dos riscos oferecidos pela utilização de tecnologias baseadas em inteligência artificial a partir de uma investigação da literatura. Por fim, o texto discorrerá a respeito das principais questões regulatórias relacionadas aos sistemas de inteligência artificial. Inicia-se a seção com a exposição de conceitos elementares de teoria regulatória e sua relação com a temática da inteligência artificial, para então passar à exposição dos pontos mais relevantes das legislações propostas na União Europeia e no Brasil.

Palavras-chave: inteligência artificial; direito da inteligência artificial; regulação da inteligência artificial; riscos da inteligência artificial; PL 2338/2023; EU AI Act

ABSTRACT

This work discusses the impacts and respective risks arising from the use of technologies based on artificial intelligence, and then addresses regulatory issues that depend on the assessment of these risks. To achieve this, firstly, common misconceptions regarding the subject are dispersed in order to establish precise definitions about artificial intelligence. From there, a typology of the risks offered by the use of technologies based on artificial intelligence is presented based on an investigation of the literature. Finally, the text will discuss the main regulatory issues related to artificial intelligence systems. The section begins with the exposition of elementary concepts of regulatory theory and their relationship with the theme of artificial intelligence, and then moves on to the exposition of the most relevant points of the legislation proposed in the European Union and in Brazil.

Keywords: artificial intelligence; artificial intelligence law; regulation of artificial intelligence; risks of artificial intelligence; PL 2338/2023; EU AI Act

SUMÁRIO

1 INTRODUÇÃO 8

2 CONCEITO DE INTELIGÊNCIA ARTIFICIAL 12

3 RISCOS DA INTELIGÊNCIA ARTIFICIAL 18

3.1 Dos riscos de responsabilização 21

3.2 Dos riscos de operação não pretendida 22

3.3 Dos riscos de compreensão 24

3.4 Dos riscos a direitos fundamentais 25

3.4.1 Riscos de violação da cidadania e da equidade 25

3.4.2 Riscos de manipulação 26

3.4.3 Riscos à incolumidade física e riscos denominados “existenciais” 27

3.4.4 Riscos referentes ao direito ao trabalho 28

3.4.5 Riscos da utilização da inteligência artificial na administração da justiça 29

4 REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL 37

4.1 AI ACT do Parlamento Europeu 44

4.1.1 Banimento de aplicações com risco de serem utilizadas para desinformação e manipulação 45

4.1.2 Banimento de aplicações com risco de serem utilizadas para “social scoring” e sistemas de identificação biométrica para o combate ao crime 45

4.1.3 Aplicações de alto risco, mas não banidas 46

4.1.4 Obrigações gerais quanto à transparência 48

4.1.5 Obrigações específicas para aplicações de alto risco 49

4.2 Projeto de Lei 2338/2023 50

5 DISCUSSÃO 53

REFERÊNCIAS 55

1 INTRODUÇÃO

As condições objetivas da vida humana em sociedade, como a disponibilidade de alimentos ou transformações religiosas e morais, influenciam e moldam as normas sociais e suas formas de aplicação. Isso faz com que os avanços tecnológicos que impactam nas formas de interação social tenham tido, em todos os momentos históricos, grande repercussão na organização jurídica das sociedades.

A revolução digital e das telecomunicações, iniciada nas universidades com seus imensos computadores, passando pela World-Wide Web (Berners-Lee *et al.*, 1992) e que décadas depois resultaram na conexão da maioria da população do planeta por meio de dispositivos móveis, impactou todos os aspectos do funcionamento da sociedade (Havick, 2000; Broens *et al.*, 2017). Por óbvio, o impacto inclui o Direito em todas as suas dimensões — em seu escopo, em suas prioridades e em sua realização material.

Uma ferramenta tecnológica amplamente utilizada pelos profissionais do direito é a consulta à legislação e à jurisprudência de forma digital. Para todos os efeitos, os antigos livros de referência foram substituídos por bancos de dados com normas sempre atualizadas e que podem ser acessados por meio de sistemas informatizados de busca. A economia não é apenas de papel e recursos físicos: a busca por informação em livros ou mesmo numa biblioteca poderia demorar horas ou dias. Com as ferramentas digitais, juristas que começaram suas carreiras antes dos anos 1990 e que lidavam apenas com textos em papel passam a lidar quase que exclusivamente com informações dispostas numa tela de computador ou mesmo, para praticidade, na palma da mão, por meio de um celular ou *tablet*.

Essa mudança também altera fundamentalmente as capacidades exigidas de um jurista. No começo do século XX, o trabalho do jurista era mais dependente de sua habilidade de memória, mesmo que para conhecer as referências úteis em termos de principais julgados e instrumentos legislativos, dado que não havia forma simples e rápida de acessá-los. Não desaparece, com o avanço das tecnologias de informação, a necessidade de conhecer com profundidade o ordenamento jurídico — e, portanto, ter na memória julgados e legislações. O foco, porém, foi alterado para a capacidade dos juristas de conjugar o que é aprendido e conhecido com a capacidade de utilizar os sistemas informatizados de busca.

Outra mudança relevante para a atuação de juristas e operadores do Direito em geral é a digitalização dos autos de processos judiciais. Uma das primeiras consequências da digitalização é a possibilidade de consulta remota aos autos de processos: o precursor da agora ubíqua tecnologia de consulta digital aos autos foi, em 1991, a implementação no STJ da possibilidade de acesso por meio de rede de computadores a andamento de processos por meio da Rede Nacional de Pacotes — Renpac —, da Embratel. Ainda na seara do processo digital, em 1998 o STJ lançava o sistema *push*, isto é, um sistema de aviso, por meio de e-mail, para informar advogados a respeito de atualizações ou modificações em processos a que estivessem vinculados (Superior Tribunal de Justiça, [20--]).

A Lei nº 11.419, de 19 de dezembro de 2006, por sua vez, tratou ineditamente da informatização do processo judicial. A lei possibilitou, em seu art. 2º, "o envio de petições, de recursos e a prática de atos processuais em geral por meio eletrônico" (Brasil, 2006), e vinculava o envio digital à assinatura eletrônica e o credenciamento prévio mediante o Poder Judiciário.

Desde o advento da internet, a transmissão digital de dados e imagens permite interações complexas por meio virtual. Os contratos eletrônicos, por exemplo, que aqueles cuja celebração depende de um sistema informático (Pinheiro, 2016), crescem cada vez mais. Além disso, as reuniões virtuais passaram a ser cada vez mais frequentes, tanto entre equipes jurídicas quanto entre advogados e clientes. O corolário dessas possibilidades é a inevitável popularização das audiências virtuais, que economiza tempo e barateia o acesso à justiça, já que se evita os custos da utilização de meios de transporte, em especial para grandes distâncias.

Algumas tecnologias, como as mencionadas acima, estão bem estabelecidas na prática jurídica, inclusive com várias delas contendo legislação para a regulamentação do seu uso. Outras tecnologias, porém, são emergentes, o que significa que apenas recentemente estão sendo reconhecidas ou utilizadas.

Uma delas é a tecnologia *blockchain*, que consiste num registro de dados disponível e atualizado a todos os usuários que participam daquela rede — o que serve, na prática, como um livro-razão contábil (Ribeiro; Mendizabal, 2021). Na modalidade "proof-of-work", ou prova de trabalho, cada bloco é composto por um dado (que é o que se quer registrar), um *hash* (ou número criptografado identificador que é obtido a partir de um cálculo que envolve o dado daquele bloco) e um *hash* do bloco anterior (Frankenfield, 2023). Como cada bloco registra o *hash* do bloco anterior e como toda a cadeia é validada por todos os usuários após cada

transação, caso alguém tente alterar um dado de um único bloco, toda cadeia será invalidada. Ainda que existam outras modalidades de implementação de *blockchain*, o exemplo em tela serve para expressar que essa tecnologia consegue, de forma descentralizada, fazer a validação de qualquer tipo de transação.

No âmbito do Direito, em geral, essa tecnologia emergente poderá ser útil para ser o substrato tecnológico para meios de troca, como moedas, para o registro de contratos, processos judiciais, escrituras de bens, direitos autorais e tantos outros.

Quanto ao Direito probatório, o campo do *e-discovery* já se impõe como uma necessidade. Basicamente, *e-discovery* refere-se à possibilidade de busca, pesquisa e recuperação de dados contidos em sistemas eletrônicos com objetivo de utilização forense, para subsidiar a formação de prova perante juízo (Oard *et al.*, 2010).

Por fim, para citar mais um exemplo de tecnologia que ganha espaço, tem-se as plataformas digitais de acordos judiciais e extrajudiciais. De fato, o Conselho Nacional de Justiça já se pronunciou a respeito da necessidade de incentivar a autocomposição:

> Visa estimular a comunidade a resolver seus conflitos sem necessidade de processo judicial, mediante conciliação, mediação e arbitragem. Abrange também parcerias entre os Poderes a fim de evitar potenciais causas judiciais e destravar controvérsias existentes (Conselho Nacional de Justiça, 2020).

Seria um enorme ganho para o Judiciário e para a sociedade se menos lides efetivamente tornassem-se processos, e a tecnologia possibilitaria, como nunca, a arbitragem ou a resolução desses conflitos já na fase de conhecimento do processo.

Das tecnologias emergentes, porém, é a inteligência artificial, por seu potencial disruptivo, que mais chama a atenção, que consiste na capacidade de máquinas realizarem tarefas complexas, cuja resolução usualmente é associada à capacidade intelectual do ser humano (Fetzer, 1990). São as tecnologias derivadas da aplicação tecnológica da inteligência artificial que compõem o cerne do escopo desta monografia.

O leitor notará que, enquanto este trabalho busca refletir a respeito de implicações materiais, riscos e reações regulatórias na sociedade em geral, para méritos de exemplificação foram priorizados exemplos voltados às profissões jurídicas.

A partir daqui o trabalho se organizará apresentando, no segundo tópico, os conceitos fundantes para o campo da inteligência artificial. No tópico seguinte, serão explorados os impactos e respectivos riscos das aplicações baseadas em inteligência artificial. No quarto tópico, serão consideradas as mais recentes reações regulatórias referentes à inteligência

artificial, abordando propostas legislativas do Brasil e da União Europeia. Por fim, a monografia encerra-se com uma discussão.

2 CONCEITO DE INTELIGÊNCIA ARTIFICIAL

Inteligência artificial, como já mencionado, é o termo utilizado para designar a capacidade de máquinas realizarem tarefas complexas, cuja resolução usualmente é associada à capacidade intelectual do ser humano (Fetzer, 1990). Há de se destacar que a expressão "inteligência artificial" não sugere que computadores efetivamente sejam inteligentes, que "pensem" ou "compreendam" como um ser humano faz. As "decisões" tomadas por um computador são meramente *outputs*, ou saídas, de um sistema informatizado, normalmente definidos por uma sequência de operações matemáticas e estatísticas. É o ser humano, e apenas o ser humano, que atribui o significado de "pensamento", ou de "inteligência", às operações realizadas pelo computador. A máquina, assim como uma pedra, um rio ou uma cadeira, não pode implicar significado, ou intencionalidade, a qualquer ação.

A questão da atribuição de significado por meio da inteligência é relevante para a discussão. Para analisar esse ponto mais a fundo evoca-se um famoso experimento mental elaborado por Searle: o quarto chinês (Searle, 1980). Searle propôs imaginar-se uma situação em que há um quarto com uma pessoa dentro, e, do lado de fora, há outras pessoas. As pessoas do lado de fora são falantes de chinês e conseguem, por meio de uma fresta, inserir folhas de papel com caracteres em língua chinesa dentro do quarto.

A pessoa dentro do quarto não sabe falar, ler ou compreender a língua chinesa. Ela tem, porém, um livro de instruções em que ela pode se basear para dar uma resposta: ao receber um carácter chinês, a pessoa dentro da sala abrirá o livro e, seguindo as instruções em português, enviará para fora do quarto, pela fresta, um ou mais caracteres chineses, que correspondem a uma resposta inteligível. Por exemplo, numa situação, o livro poderia dizer assim: ao receber um carácter com duas linhas curvas e duas linhas retas paralelas, você deve responder com o carácter que contém uma linha reta e três linhas curvas.

Esse procedimento pode se repetir de forma que as pessoas do lado de fora escrevam mensagens, as insiram no quarto e recebam respostas inteligíveis e que fazem sentido em chinês. Ao longo de diversas iterações, essas pessoas não conseguiriam descobrir que a pessoa dentro do quarto não tem qualquer compreensão da língua chinesa — para todos os efeitos, o quarto chinês consegue simular comunicação em chinês, apesar de seu operador simplesmente estar seguindo um livro de instruções.

O ponto do experimento mental é que a pessoa dentro do quarto chinês é comparável a um computador: a pessoa tem acesso a um livro de instruções, assim como o computador tem acesso a um código, que expressa um algoritmo. A pessoa, com seu livro de instruções, realiza operações meramente baseadas em aspectos formais, ou seja, nos formatos dos caracteres recebidos, a ordem em que aparecem, mas nunca em seu significado, pois os significados dos caracteres chineses são completamente indecifráveis para aquela pessoa dentro do quarto — isto é, linguisticamente, diz-se que a pessoa dentro do quarto chinês opera no nível da sintaxe, mas não no nível da semântica, pois as suas respostas não são baseadas no significado das instruções, apenas na forma dos signos e na ordem em que aparecem (Searle, 1980). O computador também opera apenas no nível da sintaxe, manipulando, como uma máquina de Turing, zeros e uns a partir de um programa, e nunca no nível da semântica, já que seu funcionamento não permitiria qualquer coisa diferente disso.

Toda tecnologia de inteligência artificial é produzida no substrato computacional que utilizamos em nosso dia a dia — são programas manipulando dados a partir de regras sintáticas. De fato, a inteligência artificial requer programas de elevada complexidade e com fatores probabilísticos, mas ainda assim são programas que seguem regras determinadas. No mesmo sentido, alguns programas de inteligência artificial, em especial aqueles identificados como "modelos de linguagem", como o ChatGPT, são chamados de "papagaios estocásticos" (Bender *et al.*, 2021). A ideia é que papagaios repetem sons ou palavras ditas a eles sem atribuir significado, e, de fato, é isso que os modelos de linguagem fazem: esses programas de computador são "treinados" a partir do processamento de uma base de dados estupenda — por exemplo, o ChatGPT (GPT-3.5) processou mais de 300 bilhões de palavras retiradas da internet a partir de sites, livros, revista. Esses programas de modelos de linguagem preveem, de forma probabilística, por meio de cálculos matemáticos, qual é o carácter ou a palavra que deveria aparecer em sequência.

Não há, de fato, real inteligência em operação, mas basicamente cálculos que projetam os próximos caracteres de uma sequência. É elementar que, do ponto de vista da implementação e da engenharia, devido à quantidade enorme de cálculos envolvidos, isso só pode ser feito, atualmente por computadores (Sloman, 2002), mas seria teoricamente possível — não do ponto de vista da engenharia ou na prática, note-se — que um ser humano realizasse esses cálculos conforme um programa com um papel e caneta por si mesmo, e obtivesse como resultado uma

sequência de caracteres. O paralelo com o quarto chinês acima apresentado é impressionantemente forte.

Ou seja, a partir da manipulação de aspectos puramente formais — da sintaxe — poderia se chegar em respostas como as dadas pelo ChatGPT. De fato, é exatamente isso que o ChatGPT faz e, portanto, o ChatGPT não "pensa" e não compreende e nunca poderá compreender semanticamente os textos de entrada e de saída — o programa apenas manipula zeros e uns (Bender *et al.*, 2021).

Toda a linguagem associada à inteligência artificial está carregada de significados que contêm paralelos com a inteligência humana. Por exemplo, a máquina seria "treinada" e "aprenderia". As unidades de programação básica de uma "rede neural" (em comparação ao sistema nervoso humano) é um "neurônio" (Hinton, 1992). Por simplicidade, usaremos a nomenclatura vigente, tendo o cuidado de entender que não há equivalência, como já demonstrado acima, entre o processo humano que chamamos "raciocínio" e as operações que máquinas realizam (Bender *et al.*, 2021).

Ao esclarecer as limitações inerentes aos programas de computador — e, por conseguinte, à tecnologia da inteligência artificial — não se está subestimando a dimensão do enorme impacto que a IA tem e terá em todos os campos da sociedade. Afinal, é pelo fato de esperar-se um enorme impacto da IA em todo o setor produtivo e na vida cotidiana que este trabalho existe. O que se busca aqui é meramente ter clareza do que significa a expressão inteligência artificial, para não incorrer em imprecisões ou mesmo confusões. Por exemplo, com um conhecimento errôneo das características substantivas inteligência artificial e como essas características diferenciam-se da inteligência humana, seria possível cair em diversos erros, como o de atribuir personalidade a uma máquina (Cosmo, 2022).

O fato de um ser humano eventualmente considerar uma máquina como uma pessoa é uma ilusão gerada pela tendência do ser humano atribuir significados ao que existe no seu meio (Bender *et al.*, 2021). O resultado das operações estatísticas e matemáticas promovidas por um modelo de linguagem como o ChatGPT é interpretado pelo ser humano que lê a sequência de caracteres gerada pelo sistema e que dá sentido ao resultado — mas o sistema em si, como já detalhado anteriormente, não tem qualquer propriedade que permite compreender semanticamente suas saídas, ou *outputs*. O sistema, portanto, não tem e não pode ter personalidade — e atribuir personalidade a um sistema de inteligência artificial é o mesmo que atribuir personalidade a uma pedra, a uma xícara, a uma calculadora de padaria ou a qualquer

outro objeto. Essa constatação, apesar de poder parecer óbvia para a maioria das pessoas, tem relevância jurídica, pois, por exemplo, só se pode responsabilizar quem tem personalidade.

Quanto à sua capacidade, a inteligência artificial pode ser categorizada em duas classes. De um lado, temos a inteligência artificial geral, ou forte, que descreve máquinas seriam capazes de efetivamente pensar, ou raciocinar, assim como faz um ser humano (Flowers, 2019). A inteligência artificial forte é um construto teórico e, apesar de existirem estudiosos que acreditam na possibilidade de sua existência, há muitos outros que não (Braga; Logan, 2017). O autor desta monografia se posiciona neste segundo grupo, em especial em consideração aos argumentos dispostos acima — de que as máquinas, ao manipular símbolos sem atribuir significado a eles, realizam um processo que é fundamentalmente distinto do que faz um ser humano ao pensar. Mesmo aqueles que acreditam na possibilidade entendem que a humanidade, no mínimo, está muito longe de conceber, do ponto de vista da engenharia, como seria construir tal tipo de sistema (Braga; Logan, 2017).

Mais importante é a segunda categoria, denominada inteligência artificial estreita, ou fraca, que se refere a todos os programas que conhecemos e que usualmente chamamos de inteligência artificial (Flowers, 2019). Trata-se das máquinas que conseguem realizar certas operações, geralmente complexas, a partir de um algoritmo — conjunto de instruções sequenciais e condicionais que uma máquina obedece, não por interpretar abstratamente o mesmo algoritmo, conforme já mencionamos, mas porque o programa fisicamente faz o processador dar um *output* desejado — e que conseguem progredir em resultados ao longo do tempo, a partir de mecanismos de "treino" e "retroalimentação".

Quando o algoritmo prepara a máquina para receber várias entradas, ou *inputs*, e modificar seus procedimentos a partir deles, obtendo novos resultados, normalmente a partir de milhares ou milhões de iterações, tem-se o que se chama de "treinamento de máquina", em oposição a um outro tipo de algoritmo que delimita mais estreitamente o que a máquina deve performar.

É perceptível, a partir da exposição acima, que existem diferenças fundamentais entre como um ser humano e como uma máquina resolvem um mesmo problema. Vale ressaltar que máquinas, em especial computadores, são instrumentos construídos por seres humanos para facilitar a resolução de problemas.

De início, computadores foram utilizados para a realização de cálculos. A máquina, neste caso, produz resultados mais céleres do que o ser humano, apesar de não tratar

semanticamente as informações processadas. Ao longo do tempo, computadores foram utilizados para aplicações mais e mais abstratas. Por exemplo, desde o fim dos anos 90 é possível dizer que seres humanos, mesmo o melhor de todos, não podem mais vencer computadores num jogo de xadrez com utilização plena de recursos de tempo e de memória.

As máquinas de xadrez nos anos 90 baseavam-se tanto em jogadas ensinadas por seres humanos — retiradas de livros de teoria enxadrística, por exemplo, e diretamente programados no computador — quanto em força bruta computacional — cálculos realizados no decorrer do jogo. Não se tratava, portanto, de uma tecnologia de inteligência artificial. O computador *Deep Blue*, por exemplo, conseguia calcular 200 milhões de posições distintas por segundo (Goodrich, 2021).

Já o ser humano faz pouquíssimos cálculos, diretamente, pois busca soluções de forma muito distinta daquela usada por máquinas. É dito, numa divertida história apócrifa, que Richard Réti — ou pode ter sido José Raúl Capablanca — um dos melhores, mais competentes e mais criativos enxadristas de todos os tempos, ao ser perguntado quantas jogadas ele calculava a cada movimento, teria respondido: “Eu vejo apenas uma jogada à frente, mas é sempre a correta.” (Winter, 2022) Apesar da dúvida a respeito da existência dessa fala, a história carrega uma perspectiva interessante: de que o raciocínio humano é carregado de intuição, criatividade e aspectos que são completamente estranhos à forma que máquinas operam. Hoje os computadores têm procedimentos mais aprimorados, com inteligência artificial, mas que correspondem, em última instância, a cálculos matemáticos e estatísticos, desprovidos de significado fora da tarefa muito específica dada pelas regras do algoritmo e pelos *inputs*.

Percebe-se que o potencial da inteligência artificial, em especial no sentido de resolver problemas complexos, buscar e organizar informação de forma efetiva e realizar análises iniciais sobre o curso a se seguir num determinado problema é de imensa utilidade em todos os campos da vida social.

Atualmente, computadores podem apenas ser aplicados para resolver problemas pontuais, conforme já visto acima, pois trata-se de inteligências artificiais fracas. Como dito, máquinas não “aprendem”, ou “pensam”, ou “conhecem”, mas apenas manipulam símbolos sem atribuição de significado a eles. Apesar de alcançarem resultados surpreendentes na resolução de certos problemas, não podem transferir o que “aprendem” para outros campos do conhecimento, nem podem demonstrar criatividade, compaixão, prudência e tantas outras

características próprias do ser humano (Braga; Logan, 2017) e que são extremamente importantes em diversas áreas, como na administração da justiça ou no provimento da saúde.

3 RISCOS DA INTELIGÊNCIA ARTIFICIAL

São muitos os benefícios que as aplicações de inteligência artificial podem oferecer — e pode-se argumentar que muitos deles sequer são conhecidos ou foram descobertos. Da mesma forma, diversas incertezas surgem com a incidência de riscos decorrentes das aplicações de inteligência artificial.

Para compreender os riscos e sua natureza, pode-se recorrer a uma tipologia, ou a pesquisas que buscam identificar os riscos já indicados na literatura especializada.

Nesse sentido, a pesquisa mais abrangente encontrada foi aquela realizada por Teixeira *et al.* (2017), em que se pesquisaram, na base *Web of Science,* artigos de janeiro de 1999 a maio de 2020 que tratassem a respeito de riscos éticos e tecnológicos associados ao uso da inteligência artificial — considerando também as expressões correlatas "aprendizado de máquina" e "orientado a dados". Foram encontrados 412 artigos para a triagem. Desses, apenas 107 foram considerados relevantes, e, por razões de língua ou acessibilidade, 104 artigos foram considerados para análise. As áreas gerais em que esses artigos estão inseridos são Ciência da Computação, Engenharia, Outros Tópicos de Ciências Sociais, Filosofia e Medicina Geral e Interna.

A partir desses artigos, foram extraídos os riscos e vulnerabilidades identificadas. Sumariza-se a seguir, na tabela 1, conforme a definição dos autores, todos os vinte e quatro riscos encontrados.

No mesmo trabalho, Teixeira *et al.* (2017) prosseguiram a uma investigação da percepção de cada um dos riscos elencados. Os achados mostram que o risco mais apontado como muito severo é aquele associado à falta de equidade, isto é, risco de tratamento parcial ou injusto, que assim foi considerado por 75% dos participantes da pesquisa. Em seguida, os riscos percebidos como de maior severidade são aqueles de víes e de proteção/privacidade de dados. Os riscos considerados menos severos são os de extinção e aqueles relacionados à semântica.

Tabela 1 — Riscos e vulnerabilidades identificadas na pesquisa bibliográfica

Conceito	Descrição
Viés	Um erro sistemático, isto é, uma tendência a aprender de forma consistentemente errada.
Explicabilidade	Qualquer ação ou procedimento realizado por um modelo com a intenção de esclarecer ou detalhar suas funções internas.
Completude	Descrição da operação de um sistema de maneira precisa.
Interpretabilidade	Descrição dos componentes internos de um sistema de uma forma que seja compreensível para humanos.
Precisão	A avaliação de quão frequentemente um sistema executa previsões corretas.
Segurança (*security*)	Implicações da utilização da IA em armas para defesa (a integração de capacidades baseadas em IA nos domínios terrestre, aéreo, naval e espacial pode afetar as operações de armas de forma combinada).
Proteção	"Lacunas" que surgem ao longo do processo de desenvolvimento em que condições normais para uma especificação completa da funcionalidade pretendida e da responsabilidade moral não estão presentes.
Semântica	Diferença entre as intenções implícitas de funcionalidade do sistema e a especificação explícita e concreta usada para construir o sistema.
Responsabilidade moral	A diferença entre um ator humano estar envolvido na causalidade de um resultado e existir o tipo de controle robusto que estabelece responsabilidade moral pelo resultado.
Responsabilização civil	Quando causar dano a terceiros, o risco de que as perdas causadas pelo dano serão arcadas pelas próprias vítimas e não pelo fabricante, operadores ou usuários do sistema, conforme apropriado.
Proteção/Privacidade de Dados	Canal vulnerável pelo qual informações pessoais podem ser acessadas. O usuário pode querer que seus dados pessoais sejam mantidos em sigilo.
Qualidade dos dados	A qualidade dos dados é a medida de quão adequado é um conjunto de dados para servir a seus propósitos específicos.
Moral	O risco de que os humanos sentirão menos responsabilidade moral em relação a decisões graves, de vida ou morte, à medida em que se aumenta a autonomia das máquinas.
Poder	A influência política e a vantagem competitiva obtida por ter tecnologia.
Sistêmico	Aspectos éticos das atitudes das pessoas em relação à IA e, por outro lado, problemas associado à própria IA.
Segurança (*safety*)	Conjunto de ações e recursos utilizados para proteger algo ou alguém.
Confiabilidade	Confiabilidade é definida como a probabilidade de o sistema funcionar satisfatoriamente por um determinado período de tempo sob condições determinadas.
Equidade	Tratamento imparcial e justo, sem favoritismo ou discriminação.
Opacidade	Provém da incompatibilidade entre a otimização matemática em alta dimensionalidade, característica do aprendizado de máquina, e as demandas de raciocínio em escala humana e estilos de interpretação semântica.
Diluição de direitos	Uma possível consequência do autointeresse na geração de diretrizes éticas para o uso da IA.
Manipulação	A previsibilidade do protocolo de comportamento em IA, particularmente em algumas aplicações, pode atuar como um incentivo para manipular esses sistemas.
Transparência	A qualidade ou estado de ser transparente.
Extinção	Risco para a existência da humanidade.
Responsabilidade	A capacidade de determinar se uma decisão foi tomada de acordo com padrões processuais e substantivos e responsabilizar alguém possível se esses padrões não forem respeitados.

Fonte: Teixeira *et al.* (2017)

A partir dos riscos evidenciados na revisão bibliográfica mencionada, propõe-se, neste trabalho, uma categorização, agrupando-se os riscos em diferentes grupos em que há fatores comuns.

As quatro categorias propostas são: (1) riscos de responsabilização: são riscos relacionados à dificuldade de responsabilização por erros ou danos cometidos por aplicações de IA; (2) riscos de operação não pretendida: são riscos relacionados à operação não pretendida pelos desenvolvedores da aplicação de IA; (3) riscos de compreensão: são riscos relacionados à falta de compreensão dos desenvolvedores ou dos usuários quanto à operação da IA; e (4)

riscos a direitos fundamentais: são riscos relacionados a ameaças a direitos fundamentais, como vida, incolumidade física, privacidade, patrimônio e trabalho.

A cada uma das categorias foi atribuída um conjunto de riscos, conforme a tabela 2 abaixo.

A forma mais adequada de compreender a tipologia proposta é no sentido de aplicação de diferentes níveis de análise para um mesmo fenômeno, e não como tipos concorrentes e mutuamente excludentes. Dessa forma, há situações, ou fenômenos, em que apenas um tipo de risco é relevante para a análise. Há outras, por outro lado, em que mais de um — ou todos — os tipos de riscos serão necessariamente empregados para sua compreensão.

Em outras palavras, ao se considerar uma situação fática, no mundo real, mais de um risco — e tipos de risco — pode ser identificado simultaneamente. Por exemplo, uma situação pode representar um risco de viés — ou seja, risco de operação não pretendida — e também um risco à proteção/privacidade de dados — isto é, a direito fundamental. De fato, nada impede que um mesmo evento carregue riscos e repercussões do ponto de vista da responsabilização, da operação não pretendida, da compreensão e dos direitos. Todas as dimensões devem sempre ser empregadas para a compreensão integral do fenômeno, mas é comum que uma ou outra sobressaia-se, a depender da situação.

Vale ressaltar que alguns riscos foram considerados pertinentes a mais de uma categoria, conforme indicado.

Tabela 2 — Categorias propostas de riscos da utilização da inteligência artificial com base nos achados de Teixeira *et al.* (2017)

Riscos de responsabilização	**Riscos de operação não pretendida**	**Riscos de compreensão**	**Riscos a direitos fundamentais**
Responsabilidade moral	Viés	Explicabilidade	Segurança (*security*)
Responsabilização civil	Completude		Proteção/Privacidade de Dados
Proteção		Interpretabilidade	Moral
Responsabilidade	Precisão	Opacidade	Poder
	Semântica	Transparência	Sistêmico
	Qualidade dos dados		Segurança (*safety*)
	Confiabilidade		Equidade
	Manipulação		Diluição de direitos
			Extinção

3.1 Dos riscos de responsabilização

Suponha-se um carro pilotado por inteligência artificial, sem qualquer entrada ou *input* humano durante seu funcionamento. Suponha também que esse carro cause um acidente de trânsito, talvez com a perda de uma vida humana. Como se daria a responsabilização? Da empresa que construiu o carro? Da empresa que construiu o *software* que pilota o carro? O dono do carro teria algum grau de responsabilização? Essa é uma questão jurídica que já está nos tribunais dos Estados Unidos (Maliha; Parikh, 2022).

Outro problema de difícil resolução, que tem interface significativa com a responsabilidade civil, mas apresenta diferenças importantes, é quando um sistema operado por inteligência artificial precisa escolher entre dois cenários em que há prejuízo à incolumidade física de usuários do sistema ou de terceiros afetados. Ainda no exemplo proposto acima, se o carro pilotado por inteligência artificial não tiver tempo de realizar frenagem, mas necessariamente, tenha que "decidir" entre atingir uma pessoa ou atingir outra pessoa; ou ainda, se o sistema operado por inteligência artificial precisar escolher entre proteger a vida de seu usuário às custas da vida de um pedestre ou outra pessoa não diretamente envolvida — estes são casos em que há escolhas de ordem ética em que o Direito provavelmente precisará tomar uma posição prévia por meio da regulação específica.

De fato, esta é uma categoria de problema muito similar ao dilema do bonde ("*trolley problem*") proposto na década de 60 por Philippa Foot (Thomson, 1985). Neste clássico problema de Ética Aplicada, realiza-se um experimento mental em que um operador de um bonde se vê num dilema: o bonde, se não desviar de seu trilho atual, atropelaria cinco pessoas; o operador pode mudar o bonde de trilhos, o que resultaria em "apenas" uma outra pessoa sendo atropelada, ficando as cinco pessoas a salvo; e o operador do bonde não consegue parar o veículo a tempo, o que significa que ele tem, efetivamente, que escolher entre a inação ou a mudança de trilhos.

O que o operador do bonde deve fazer? Ele tem a obrigação ética de fazer ou de não fazer algo nessa situação? Muitos são os debates em torno desse exercício mental — e não se pretende, neste trabalho, entrar no mérito da questão em específico, ou nas soluções encontradas pela deontologia ou pelo utilitarismo. Um caminho possível em casos como esses é optar por uma abordagem regulação mínima, que contenha um conjunto de comportamentos inaceitáveis no uso de aplicações de inteligência artificial (Page *et al.*, 2018).

A questão é que os legisladores e os operadores do Direito, a partir do uso ubíquo das tecnologias de inteligência artificial, terão que se deparar, eventualmente, com situações que abstratamente se encontram na alçada do dilema do bonde, mesmo que em outros campos, como na Medicina. Quando o assunto é inteligência artificial, esses e outros desafios aguardam as sociedades do presente e do futuro.

3.2 Dos riscos de operação não pretendida

Existe o risco de situações em que os processos estatísticos e decisórios das máquinas resultem em danos para usuários ou para terceiros. Este é um impacto relacionado ao risco de viés de máquina ("*machine bias"*, em inglês), que é a tendência de um modelo de aprendizado de máquina em fazer previsões incorretas ou injustas por conta de erros sistêmicos no modelo em questão ou nos dados utilizados para treiná-lo. Normalmente, o termo "viés" é utilizado para descrever resultados sistemicamente enviesados na operação de um sistema de inteligência artificial que resulte em discriminação por raça, sexo, religião, etnia, nacionalidade, posição social ou qualquer outro atributo humano, bem como qualquer informação factualmente errada atribuída a algo ou alguém (Rouse, 2023).

Para melhor situar o exposto acima, é importante destacar uma definição: aprendizado de máquina, ou *machine learning*, é uma forma de construir *software* em que o programador não codifica direta e explicitamente a sequência lógica de operações que resultará nos *outputs*, ou saídas, do programa, mas cria parâmetros para que o programa, a partir de um número de iterações, "aprenda" e comece a fornecer *outputs* desejados (Tan, 2017). Na clássica definição de Mitchell (1997), "Diz-se que um programa de computador aprende com a experiência E em relação à classe de tarefas T e à medida de desempenho P, se seu desempenho nas tarefas medidas por P melhora com a experiência E".

Apesar do nome, não se trata propriamente de "aprender" como um ser humano, que é algo que sempre incluirá algum grau de compreensão de significado — conforme já discutido, os programas de inteligência artificial consistem meramente em operações matemáticas e estatísticas que ocorrem num substrato de circuitos eletrônicos. Como as máquinas trabalham apenas no nível da sintaxe, e nunca da semântica, o "conhecimento" que elas obtêm não é generalizável, não podendo ser aplicado fora da estreita faixa das relações sintáticas que elas

conseguem manipular — trata-se, portanto, de inteligência artificial estreita, ou fraca. Utilizaremos a nomenclatura vigente, porém, por simplicidade.

O viés de máquina está altamente correlacionado com os dados que alimentam o aprendizado da máquina. Os conjuntos de dados utilizados normalmente são enormes, de forma que é impossível para um ser humano — ou um time — rever todas as informações inseridas, bem como prever todas as possíveis consequências e interações que os dados inseridos podem ter no sistema de inteligência artificial.

Sabe-se que, enquanto normalmente atribui-se os vieses a bases de dados tendenciosas (Cozman; Kaufman, 2022), há diversas outras fontes de vieses já identificadas, como, por exemplo, viés na geração dos dados e viés nas escolhas realizadas por desenvolvedores e programadores. Há também uma tipologia de vieses nas bases de dados, que incluem viés no processo de rotulagem dos dados e viés nos dados de treinamento dos algoritmos (Cozman; Kaufman, 2022). Portanto, a insegurança em relação a vieses de máquinas, e grande falta de controle em relação aos *outputs* das máquinas, requerem que o uso da inteligência artificial sempre se dê sob supervisão humana, garantidas interfaces para que o gestor do processo possa efetivamente intervir quando necessário.

Uma preocupação relevante é com sistemas de inteligência artificial que fazem a gestão de estruturas essenciais para o funcionamento da sociedade, como aquelas voltadas ao fornecimento de água, eletricidade, combustíveis, ou mesmo aquelas voltadas ao transporte em seus mais diversos modais ou mesmo às telecomunicações.

Ao se tornarem cruciais para a operação das estruturas, podem gerar dependência da sociedade em relação a eles — o que pode se provar crítico em caso de falhas por defeitos ou mesmo por ataques de hackers. Os efeitos de situações assim incluem cortes de energia, de água ou de gás, paralização do fluxo de transportes e dano às telecomunicações.

Aliás, em relação a serviços públicos em geral, e considerando que a logística moveu para uma lógica "*just in time*" em grandes centros urbanos, a maioria das grandes cidades do mundo têm estocado menos do que uma semana de suprimentos alimentícios e ainda água potável. Isso significa que uma massiva descontinuidade em serviços de transporte poderia comprometer milhões de vidas (Page *et al.*, 2018).

3.3 Dos riscos de compreensão

Considerando as situações em que os processos estatísticos e decisórios das máquinas resultem em danos para usuários ou para terceiros, existe um risco adicional: a potencial — e provável — falta de transparência em relação aos resultados obtidos pelas máquinas. Este risco remete às ideias de transparência e opacidade, que serão desenvolvidas a seguir.

O processo de geração de *outputs* de uma inteligência artificial é desconhecido para o usuário da tecnologia, de forma geral, e isso é um tanto quanto esperado. A esta falta de transparência dá-se o nome de opacidade do sistema. O que surpreende muitas pessoas é que o funcionamento exato dessas tecnologias é, em boa medida, desconhecido pelos próprios desenvolvedores — isto é, eles também precisam lidar com a opacidade dos sistemas que eles mesmos constroem (Burrell, 2016). Isso se dá por várias razões: primeiro, como já dito anteriormente, a máquina, nas tecnologias de inteligência artificial, executa programas de complexidade muito alta. Imagine que uma pessoa construa uma espécie de labirinto inclinado complexo, e que solte uma bola no topo. Apesar de o labirinto ter sido criado por um ser humano, é possível que essa pessoa não consiga dizer com precisão onde a bola irá parar, mas consiga, talvez, estimar.

Imagine um sistema bilhões de vezes mais complexo. Os programadores *a priori* não sabem qual será exatamente o *output* esperado para muitas das situações, ou *inputs*, apresentadas. Além disso, mesmo que fosse possível traçar o "processo decisório" da máquina, ainda assim seria algo muito difícil de visualizar, pois poderia incluir milhões ou bilhões de iterações distintas, com cada uma delas não tendo um significado específico ou traduzível diretamente para a linguagem humana.

Há também opacidade que é intencionalmente projetada, como por questão de segredo comercial (Hutson, 2021). Enquanto não é possível deslindar como toda operação de um sistema se dá, empresas e desenvolvedores devem ser instados a fornecer as informações necessárias e serem o mais transparentes possível, para que as pessoas compreendam como estão, de fato, interagindo com os sistemas.

3.4 Dos riscos a direitos fundamentais

"[H]á determinados direitos que são matrizes de todos os demais; são direitos sem os quais não podemos exercer muitos outros" (Salgado, 1996). Estes são os direitos fundamentais, reconhecidos pela Constituição Federal de 1988 no Título II (Brasil, 1988) como o direito à vida, à incolumidade física, à privacidade, ao trabalho e ao patrimônio.

Realiza-se abaixo, com base nos riscos identificados em Teixeira *et al.* (2017), a enumeração e análise de alguns dos riscos. Vale observar, porém, que a lista dos bens e direitos potencialmente afetados por riscos gerados pelo uso da inteligência artificial estará sempre em aberto, por força de não se conhecer precisamente as possibilidades desse tipo de tecnologia.

3.4.1 Riscos de violação da cidadania e da equidade

O funcionamento dos sistemas de inteligência artificial e sua capacidade de lidar com grandes quantidades de dados que, de outra forma, não poderiam ser analisados ou processados por seres humanos permite que entidades, como Estados e corporações, consigam traçar perfis detalhados de cidadãos ou usuários para controlar seu comportamento e restringir acesso a serviços ou benefícios. Esse tipo de classificação é chamado de "pontuação de crédito social", ou "social scoring".

O principal problema desse tipo de pontuação é o poder que a instituição detentora das informações tem sobre os indivíduos. É considerada, de forma geral, uma prática abusiva e ligada a regimes autoritários, contrariando direitos fundamentais.

Outra dimensão de interesse é relacionada à aplicação e execução da lei ("*law enforcement"*). Vieses ou mau uso da tecnologia desse sentido são uma grande ameaça aos direitos fundamentais.

É evidente que em contexto sensíveis, como na administração da justiça ou na persecução penal, resultados enviesados na produção de peças jurídicas ou na análise de casos concretos são extremamente preocupantes, pois podem comprometer a qualidade da administração da justiça em si. De fato, as formas com que os vieses de máquina podem se expressar e prejudicar pessoas em diferentes contextos são muito difíceis de identificar de antemão, de mapear e de prevenir completamente.

Na mesma esteira, um problema bastante conhecido é em relação à identificação de pessoas em imagens, com sistemas historicamente tendo maior dificuldade em identificar pessoas negras (Birhane, 2022). Sistemas que sejam usados futuramente para analisar corpos probatórios que incluam imagens poderão, se enviesados, não identificar corretamente pessoas, ou mesmo confundir diferentes pessoas.

Nos Estados Unidos já se usam programas de cálculo de riscos — os chamados *risk assessment scores* — no sistema penitenciário. Um desses programas é o *Correctional Offender Management Profiling for Alternative Sanctions* — COMPAS, que informa, com base em mais de uma centena de pontos de informação coletados a respeito do réu, se a sua probabilidade de reincidência é baixa, média ou alta. Esses sistemas são utilizados para informar o sistema judicial em suas diferentes instâncias, e influenciam, inclusive, nos valores que réus precisam pagar em suas finanças (Castro, 2019).

Não existem apenas exemplos anedóticos de previsões muito incorretas por parte desse sistema, com criminosos inveterados recebendo avaliações de risco menores do que réus primários, mas dados históricos já mostram que o sistema COMPAS falsamente identifica réus negros como futuros criminosos a uma taxa duas vezes maior do que o faz com réus brancos (Castro, 2019). Mediante essa situação, além de trabalhar para a correção dos vieses que resultam nessas injustiças, é importante comparar, de forma científica, os vieses humanos com os vieses de máquina: há aqueles que acreditam que as máquinas produzem resultados menos enviesados que os seres humanos (Atkinson, 2016), mas com a vantagem de serem mais céleres. Mais estudos são necessários para determinar a validade dessa hipótese.

3.4.2 Riscos de manipulação

Estes riscos estão ligados à possibilidade de sistemas de inteligência artificial serem utilizados para enganar e manipular pessoas. As ações maliciosas poderiam ter como objetivo causar dano direto a alguém ou a seu patrimônio, fazer marketing manipulativo, causar instabilidade política, fraudar eleições e tantas outras possibilidades nefastas.

Esse risco se encaixa na possibilidade de violação de direitos porque, na pesquisa bibliográfica, foi identificado um fator denominado “Poder”, que foi definido como “[a] influência política e a vantagem competitiva obtida por ter tecnologia.” (Teixeira *et al.* 2017). Portanto, a possibilidade de manipulação encaixa-se nessa ampla definição.

Algoritmos de inteligência artificial podem criar conteúdo enganoso e distribuir esse mesmo conteúdo para pessoas que tenham maior tendência de não perceber que estão sendo enganadas. Isso gera um sério problema na formação da opinião pública e em eleições.

Na mesma esteira, com a inteligência artificial é possível criar material audiovisual falsificado, o que se chama *deep fake.* É possível criar imagens, áudios e vídeos preparados digitalmente que mostram pessoas fazendo ou dizendo coisas que elas não fizeram.

Ainda, a exploração de vieses ou tendências pessoais para fazer marketing manipulativo é uma preocupação que envolve também as grandes empresas de redes sociais. Sabe-se que é possível utilizar aplicações de inteligência artificial para induzir pessoas a darem uma resposta determinada num questionário ou a agirem de uma forma específica (Petropoulos, 2022). A quantidade monumental de dados que empresas de tecnologia detém de seus usuários pode ser usada para utilizar técnicas de venda que inclusive podem ser consideradas viciantes (Petropoulos, 2022).

3.4.3 Riscos à incolumidade física e riscos denominados "existenciais"

Alguns pesquisadores têm a opinião de que o progresso das tecnologias baseadas em inteligência artificial representa um risco existencial para humanidade, inclusive no sentido de representar risco de extinção da espécie humana em razão das tecnologias tornarem-se incontroláveis (Stuart & Peter, 2009). Evidentemente, essa é uma hipótese bastante disputada, com diversos outros pesquisadores tendo entendimento contrário (Brooks, 2014). Além disso, entre riscos existenciais incluem-se uma eventual corrida entre potências militares para o desenvolvimento de armas baseadas em inteligência artificial (Geist, 2016).

Independentemente das opiniões a respeito de riscos existenciais da IA, não são poucas as figuras públicas do setor tecnológico que pediram algum tipo de controle em relação à inteligência artificial. Por exemplo, Sam Altman, CEO da OpenAI — que disponibiliza o ChatGPT — disse perante um subcomitê do Senado americano: "Eu acho que se esta tecnologia der errado, poderá dar muito errado [...] Nós queremos trabalhar com o governo para prevenir isso." (Kang, 2023)

Uma carta aberta assinada por Elon Musk, cofundador da OpenAI, Emad Mostaque, fundador da Stability AI e Steve Wozniak, cofundador da Apple pede expressamente pela interrupção de pesquisas com modelos de inteligência artificial mais poderosos que o ChatGPT

por pelo menos seis meses, até que se institua governança e regulação adequadas (Future of Life Institute, 2023).

Em que pese o tom catastrófico que é utilizado para descrever os riscos existenciais, a maioria das pessoas não os consideram os riscos mais severos da utilização da inteligência artificial (Teixeira *et al.*, 2017). Uma situação que requer mais imediata atenção e que tem impactos já hoje é o uso de sistemas de inteligência artificial que podem resultar em riscos para a segurança e a incolumidade física das pessoas: muitas aplicações de IA lidam com gestão de transporte, armazenamento de materiais que oferecem risco à saúde se mal manejados e até mesmo operação direta de transporte, como nos carros que são dirigidos por inteligência artificial sem intervenção humana. Discussões sobre riscos futuros são sempre proveitosas, desde que se não esqueçam dos riscos que precisam ser tratados imediatamente.

3.4.4 Riscos referentes ao direito ao trabalho

Outro impacto jurídico das tecnologias de inteligência artificial que é objeto de muito interesse é o impacto projetado no mundo do trabalho, em especial, quanto à substituição da mão de obra humana pela máquina (Brower, 2023), o que tem diversas implicações em várias áreas do Direito, como no Direito Civil, Direito do Trabalho e Direito Comercial.

Quando se pensa no uso da inteligência artificial, a ideia que se tem é, normalmente, de disrupção de processos e do mercado de trabalho, com impactos significativos em cadeias produtivas inteiras. Com a popularização e facilitação do acesso à inteligência artificial, o perfil dos trabalhos passíveis de substituição fica cada vez mais tendente àquelas atividades consideradas tipicamente intelectuais. Uma pesquisa mostra que mais de um quarto das pessoas com ensino superior nos Estados Unidos teme perder seus empregos para inteligência artificial (Morikawa, 2017). O horizonte disruptivo causa, por si mesmo, tensões e preocupações. Por exemplo, há aplicativos hoje, como o *Midjourney* e o *Stable Diffusion*, que geram imagens completamente construídas a partir de inteligência artificial. Esses aplicativos utilizam bilhões de imagens para interpolar e criar imagens inéditas. Com base nisso, já existe uma reação jurídica, uma lide em que se alega a quebra de direitos de propriedade intelectual, contra alguns desses aplicativos (Vincent, 2023).

Um aspecto impressionante a respeito de ferramentas como essas é que realizam trabalhos usualmente considerados criativos. Em realidade, as imagens geradas, apesar de

esteticamente aprazíveis e muitas vezes detalhadas, são fruto apenas de uma interpolação matemática, a aplicação de uma técnica estatística. O resultado é, portanto, plenamente dependente das entradas, ou *inputs*, dados para o computador (Butterick, 2023).

Situações como essa tenderão a se repetir no futuro, em diversas vertentes do mercado de trabalho. Aliás, vale a pena recuperar um conceito proposto pelo economista Joseph Schumpeter (1942) denominado destruição criativa. Schumpeter imagina a destruição criativa como um fenômeno econômico em que as forças produtivas de uma sociedade se reorganizam por meio da inovação e da competição mercadológica, muitas vezes de formas disruptivas e levando a alocações inesperadas de trabalho e capital, o que resulta em progresso econômico. Essa reorganização se dá, por exemplo, com a introdução de novas tecnologias, como a inteligência artificial.

Os exemplos de potencial substituição de seres humanos por inteligência artificial no mercado de trabalho são muitos. Um estudo recente, entretanto, buscou delimitar mais claramente em que partes da economia os impactos seriam mais significativos. A partir da aplicação de uma metodologia em que se compara descrições de empregos ou posições já ocupadas por seres humanos com descrições de patentes ligadas às tecnologias de inteligência artificial, é possível estimar quais carreiras estão mais expostas a esse tipo de automação. Descobriu-se que a inteligência artificial tem o maior impacto esperado em trabalho que seja altamente especializado, diferentemente do que aconteceu com a introdução de *software* e robôs ao longo do século XX, que atingiu trabalhos mais manuais, mesmo que especializados, ou menos especializados em geral (Webb, 2019).

Ainda no contexto do mercado de trabalho, é possível o uso da inteligência artificial para questões de acesso ao emprego, como recrutamento, bem como no processo decisório de desligamento. Nesses casos, o risco de vieses é muito significativo.

3.4.5 Riscos da utilização da inteligência artificial na administração da justiça

Aliás, em falar no mercado de trabalho, vale a pena pensar sobre como a inteligência artificial pode impactar os profissionais do Direito e a prática jurídica, para além daqueles riscos já considerados quanto à cidadania e à equidade, em seção anterior. De fato, se os novos produtos e resultados obtidos por meio de inteligência artificial geram ondas de impacto e de

expectativa no mercado de trabalho como um todo, é claro que a administração da justiça não poderia ficar indiferente.

No escopo deste trabalho, quando se fala em administração da justiça, deve-se pensar nas funções exercidas pelo Poder Judiciário em si, bem como naquelas funções que a Constituição Federal de 1988 define como essenciais à função jurisdicional do Estado ou à administração da justiça — aquelas exercidas pelo Ministério Público, pela Defensoria Pública e pela Advocacia, pública ou privada (Brasil, 1988). Também abordaremos as funções de suporte necessárias ao exercício dessas atividades, como de gestão administrativa ou documental.

Walker, em sua obra *On Legal AI*, uma das primeiras tentativas a tentar relacionar as possíveis aplicações da inteligência artificial no mundo jurídico, traz:

> "A realidade competitiva delineada anteriormente reforça nossos quatro axiomas, especialmente acerca do baixo custo e escalabilidade do software versus processos puramente manuais. Você precisa de plataformas integradas de IA para escalar a qualidade e nuance valorada do que você faz, até para manter seus faturamentos e fontes de receita atuais. Com uma aplicação habilidosa de IA, você pode escalar essas receitas e os lucros concomitantes. O mais importante é que você pode ajudar mais pessoas com mais frequência." (2021, local. RB-4.1)

O trecho acima mostra de forma sintética boa parte das expectativas em relação às aplicações de inteligência artificial no Direito e, mais especificamente, na administração da justiça. Walker prossegue e descreve quais são as características dos sistemas de inteligência artificial que são interessantes para as aplicações jurídicas:

> "O que torna essas tecnologias diferentes de uma roda ou limpador de para-brisa ou qualquer aparelho genérico perante o direito? Pelo menos três coisas: (1) autonomia, (2) evolução autônoma e aprendizado automático, e (3) capacidade de autorreplicação. [...]
>
> Sistemas autônomos avançados e futuros representam um desafio ao processo jurídico tradicional, e até ao raciocínio jurídico analógico tradicional. Em outras palavras, eles desafiam os meios pelos quais os juristas trabalham, não seu quadro intelectual. Esse desafio inclui velocidade e magnitude de dados." (Walker, 2021, local. RB-8.4)

Walker também pondera diretamente a respeito da substituição dos profissionais do Direito por aplicações de inteligência artificial:

> "Muitos tecnologistas jurídicos recomendam suplantar totalmente os advogados por IA e diversas plataformas de software. Na minha opinião, é uma visão de mundo insensata, mesmo quando é possível. Primeiro, há um monte de coisas boas na prática do direito. Substituir os advogados por *software* (ou *software* com consultores ou *software* com

> contadores) oferece o risco de perder um imperativo moral, assim como um conjunto de habilidades peculiares em engenharia textual. Máquinas de software e outras profissões carecem dos requisitos da ética jurídica. Segundo, essa abordagem afasta de imediato exatamente aquelas pessoas, os advogados, com os conjuntos de habilidades e experiência de negócios para fazer uma boa IA jurídica se tornar realidade." (Walker, 2021, local. RB-9.4)

Isso nos faz pensar que, em alguma medida, a tecnologia para, ao menos de forma rudimentar, tentar substituir a ação humana em processos jurídicos já existe ou está em vias de existir, o que não significa que possa fazer com a qualidade e a competência devidas. A aplicação da inteligência artificial no contexto da Administração da Justiça de forma totalmente independente apresenta riscos bastante graves, tanto pela questão da qualidade do serviço prestado, passando pelos vieses de máquina e falta de transparência já mencionados, quanto pela anulação do fator humano, essencial para a prestação jurisdicional, em especial em situações que fogem da atuação média do operador do direito.

Em especial em situações em que a parte tem *jus postulandi*, como no contexto da Justiça do Trabalho ou nos Juizados Especiais Cíveis, por exemplo, é provável que empresas fornecerão como serviço a automatização praticamente integral de boa parte das peças e atuação no processo, dados os *inputs* pela parte interessada, ou usuário. Num momento posterior, essa faculdade poderia se expandir para configurar a efetiva substituição de advogados por empresas que forneçam serviços jurídicos e que, em realidade, seriam proprietárias de sistemas de inteligência artificial capazes de atuar simultaneamente numa infinidade de casos.

Essa tendência seria, então, a aplicação da inteligência artificial para os leigos, isto é, aqueles que não são profissionais treinados nos assuntos do Direito, em especial por meio de um software que consiga responder perguntas jurídicas (Zhong *et al.*, 2020). Este é um uso que logicamente traz também diversas dificuldades. Afinal, alguém sem treinamento no Direito não tem capacidade de discernir entre dados úteis compilados por inteligência artificial e artefatos incorretos ou simplesmente inválidos. Pesquisas mostram que, atualmente, a partir de resultados experimentais, modelos de inteligência artificial não conseguem responder questões legais de forma satisfatória (Zhong *et al.*, 2020).

Apesar do tópico em tela parecer assunto de livros de ficção científica, muitas pessoas já levam essa possibilidade bastante a sério. Nos Estados Unidos, já existe uma companhia chamada DoNotPay, que promete fornecer o primeiro "advogado robô". Aliás, a companhia de fato tentou usar seu "advogado robô" ao vivo num julgamento real, mas foi impedida por uma

reação jurídica: seus administradores foram ameaçados de prisão por prática de advocacia sem licença (Allyn, 2023).

Como garantir a qualidade e a confiabilidade da atuação da inteligência artificial na advocacia, em especial frente a um usuário que normalmente não tem qualquer treinamento na técnica jurídica ou em seus pressupostos? Indo um passo além, se há aqueles que pensam em "advogado robô", também poderia se pensar num "juiz robô", que traria consigo uma gama de outras complicações. Que linha hermenêutica tal *software* deveria seguir, dentre tantas linhas que competem entre si? Como explicar para as partes de um processo a *ratio decidendi*? Considerando a opacidade de tais sistemas, que é uma questão que até seus desenvolvedores precisam lidar, como informar às partes o porquê uma linha argumentativa foi utilizada em detrimento da outra? Por critérios probabilísticos?

Nessas condições, é provável que as pessoas não aceitem ser julgadas — absolvidas ou condenadas — por uma máquina. Isso é apoiado pela noção de que seres humanos preferem o trabalho de outro ser humano, e não de uma máquina, em contextos de consumo simbólico, ou seja, quando se consome algo no sentido de expressar algo sobre suas crenças ou personalidade (Granulo *et al.*, 2020). Além disso, além da precaução causada pelos riscos já mencionados, de forma geral os seres humanos aplicam às máquinas outros vieses, tendendo a confiar menos nelas, o que é denominado "aversão algorítmica" (Dietvorst *et al.*, 2015).

Para referência, Lage (2022), além de conter uma valiosa exposição sistemática da literatura a respeito das aplicações de IA no contexto do Direito, expõe diversas aplicações já em operação ou planejadas da inteligência artificial no âmbito da administração da justiça ao redor do mundo, incluindo um projeto da Estônia de automação por meio de inteligência artificial de julgamentos que podem ser revisados por um magistrado.

Atualmente, todos os riscos acima elencados são mitigados da mesma forma: com a presença humana em todas as etapas e instâncias da administração da justiça. Enquanto as máquinas forem vistas como ferramentas para profissionais — e de fato, é o que são —, essas preocupações serão amenizadas.

Busca-se agora enumerar algumas das aplicações potenciais das tecnologias de inteligência artificial na administração da justiça, como forma de apoio aos profissionais do Direito. Por óbvio, o rol a seguir não é taxativo, mas exemplificativo.

Um primeiro ponto de utilização relevante para a inteligência artificial no contexto da administração da justiça é para a orientação jurídica dos próprios operadores do Direito por

meio do uso de corpos de dados que normalmente não poderiam ser processados por um ser humano ou por uma equipe. Por exemplo, a análise por inteligência artificial dos julgados de determinado juiz pode trazer percepções de padrões que podem direcionar a ação do causídico em determinadas situações.

Outra aplicação possível está no campo da pesquisa jurídica: as profissões ligadas ao Direito são notoriamente conhecidas por exigir uma quantidade grande de estudos ao longo de toda a vida profissional, já que o ordenamento jurídico está em constante evolução e os casos que um profissional precisa cuidar apresentam grandes variações entre si. A novidade, no Direito e nas profissões jurídicas, é a regra, e não a exceção.

Um estudo de 2019 da Bloomberg Law, ocorrido nos Estado Unidos, descobriu que 84% dos advogados acreditam que fazer as minutas para os argumentos jurídicos a serem utilizados nas peças processuais é uma das tarefas que mais consomem tempo — consumindo, em média, 20 horas de trabalho mensais — enquanto 75% acham o mesmo a respeito da tarefa de revisar e analisar os argumentos dados pela parte contrária no processo (Bloomberg Law, 2023).

Como a pesquisa jurídica é uma necessidade para todo operador do Direito, o apoio da inteligência artificial nessa seara é muito promissor para aumentar a produtividade no contexto da administração da justiça. A ferramenta de inteligência artificial traz a possibilidade da realização de pesquisas aprofundadas sobre jurisprudência, legislação ou doutrina numa fração do tempo que uma equipe jurídica completa realizaria, auxiliando tanto o processo de decisão de um magistrado como a atuação por parte do Ministério Público, Defensoria Pública e advogados.

Uma outra aplicação relevante para a inteligência artificial é, a partir do ordenamento jurídico vigente e das circunstâncias conhecidas do caso em questão, tentar prever o resultado do julgamento (Zhong *et al.*, 2020).

No contexto de sistemas *Civil Law*, isso se dá em especial por meio de um processo de subsunção dos fatos concretamente considerados às normas abstratas. Sistemas de inteligência artificial mostram-se proficientes em casos que ocorrem com maior frequência no sistema jurídico, mas tem problemas de precisão quando o caso é de baixa frequência, isto é, quando é muito diferente do que se usualmente trata em determinado sistema jurídico (Zhong *et al.*, 2020).

Em jurisdições em que o sistema jurídico se caracteriza como *Common Law*, é útil recorrer à correspondência entre casos similares, pois decisões são feitas conforme casos similares e representativos do passado. A inteligência artificial pode auxiliar o operador do Direito a encontrar com celeridade quais casos são mais similares ao objeto de seu estudo (Zhong *et al.*, 2020).

Aliás, com a progressiva redução das distinções entre sistemas *Civil Law* e *Common Law*, é de se pensar que ambas as abordagens acima mencionadas deverão ser utilizadas em qualquer jurisdição.

Aplicação de grande relevância da inteligência artificial é na execução de tarefas administrativas, como manejo e organização de documentos. Também se vislumbra que a inteligência artificial possa ajudar na gestão geral de processos, em seus tempos e momentos, bem como na organização administrativa dos órgãos da Justiça, do *parquet*, das Defensorias Públicas e dos escritórios de advocacia.

Uma pesquisa recente com advogados dos Estados Unidos mostrou que, na média, aqueles advogados gastavam 31% de seu dia de trabalho lidando com assuntos que estavam diretamente ligados aos interesses de seus clientes — ou seja, apenas 31% de seu dia de trabalho correspondia a *billable hours* (Clio, 2020).

O espaço para ganhos em produtividade, então, para que operadores do Direito possam dedicar mais tempo às atividades finalísticas é muito significativo.

Uma aplicação de inteligência artificial que apresentará grande impacto no dia a dia dos operadores do Direito é o suporte à redação de peças jurídicas. As tecnologias de inteligência artificial possibilitam a redação de peças jurídicas num tempo extremamente reduzido. Quanto mais padronizável for a peça, melhores serão os resultados e, mesmo para casos de baixa frequência, em que a peça produzida pela máquina apresenta um maior número de imperfeições, o ganho de produtividade ainda será muito significativo — tanto em um caso quanto em outro, a ideia é que o profissional iniciará a partir de um texto que precisará complementar e retocar, em vez de iniciar a partir de uma página em branco.

Outra aplicação relacionada à administração da justiça para a inteligência artificial é o tratamento e a sumarização de dados e informações (Zhong *et al.*, 2020). Tal aplicação seria útil para a redação de sumários ou resumos de casos, de peças processuais ou de contratos. Além disso, seria bastante eficiente no tratamento de quantidades muito maiores de dados,

como nos casos em que é necessário realizar análises referentes a *compliance*, como em auditorias ou processos de devida diligência.

Mencionou-se na Introdução desta monografia a realidade do *e-discovery*, que já não é novidade para o mundo jurídico, cujos procedimentos poderão ser aprimorados pela inteligência artificial. Considerando a capacidade desses sistemas de encontrar relações e conexões em grandes corpos de dados, poderá ser utilizada para potencializar a probabilidade de achados e descobertas nessa seara. Operações de *discovery* que eram realizados por dezenas ou mesmo centenas de advogados poderão ser realizadas por um número muito reduzido.

Vê-se, pelas aplicações da inteligência artificial à administração da justiça acima enumeradas, que os ganhos voltados à eficiência e à produtividade são bastante perceptíveis.

Em relação à eficiência, pensa-se no potencial de tempo e recursos financeiros a serem economizados. Quanto à eficiência, deve-se lembrar do contexto processual brasileiro. Atualmente, 80 milhões era o número de processos tramitando na Justiça brasileira em 2022, (Crepaldi; Goes, 2022) o que torna inviável para o Judiciário julgar todos eles em tempo hábil e razoável, conforme a expectativa da sociedade.

O Conselho Nacional de Justiça realizou levantamento e identificou que há 19% de déficit de juízes no Brasil (Conselho Nacional de Justiça, 2017). Entretanto, mais de 90% da população acha a Justiça brasileira muito lenta (Fundação Getulio Vargas, 2019), o que pode fazer pensar que, provavelmente, suprir o déficit apontado de 19% de juízes não parece ser suficiente: é necessário repensar boa parte da lógica organizacional do Judiciário, no sentido de atender de forma mais célere à sociedade.

Os ganhos de eficiência que a inteligência artificial traz pode ajudar a reduzir sensivelmente o hiato entre a expectativa da sociedade e os resultados que o Judiciário consegue produzir quando os gargalos estiverem relacionados à produção de minutas para decisões, compilação de informações e dados relevantes à decisão ou qualquer outra em que o processo possa ter um incremento de eficiência pelo uso de tecnologias de inteligência artificial.

Além de ganhos de eficiência e produtividade, há de se considerar os ganhos relacionados ao acesso à justiça. De forma geral, as tecnologias que promovem a redução de custos de provimento de determinado bem ou serviço podem, também, promover a redução de custos para o usuário final do serviço.

Sabe-se que custear causídicos de um processo pode ser bastante dispendioso a um ponto que, para parte significativa da população, é, hoje em dia, economicamente proibitivo.

Um advogado ou escritório de advocacia que consiga potencializar sua atuação por meio do uso inteligente das tecnologias conseguirá atender mais pessoas com mais qualidade e menor custo. Podem surgir, inclusive, nichos de mercado em que escritórios, por meio da utilização de novas tecnologias, multipliquem tanto sua eficiência que certas modalidades de serviço jurídico se tornem muito menos dispendiosas.

As tecnologias de inteligência artificial serão, de certo, grandes aliadas daqueles que se adaptarem a elas, utilizando-as em todo seu potencial. Isso é verdadeiro não apenas para os operadores do Direito, mas para praticamente todas as demais profissões.

4 REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL

Mediante o poder disruptivo e de inovação que a inteligência artificial traz consigo, bem como seus potenciais riscos, vê-se, ao redor do mundo, várias reações regulatórias aos diferentes usos desse tipo de tecnologia.

O itinerário que será seguido nesta seção é apresentar os fundamentos teóricos da regulação, para então verificar aspectos específicos da regulação de sistemas baseados em inteligência artificial. Por fim, serão feitas considerações quanto às reações regulatórias mais relevantes da atualidade.

Uma primeira categoria de definições compreende a regulação “como uma metáfora derivada de sistemas biológicos ou mecânicos e entregue às versões mais simplificadas dos mecanismos de controle utilizados para alterar o curso do sistema regulado rumo à direção desejada” (Iorio Aranha, 2019, p. 37).

Nesse sentido, a regulação “significa um processo de realimentação contínua da decisão pelos efeitos dessa decisão, reconformando a atitude do regulador em uma cadeia infinita caracterizada pelo planejamento e gerenciamento conjuntural da realidade” (Iorio Aranha, 2019, p. 38).

Ao compreender a regulação como tecnologia, uma outra definição compreenderia a regulação como “tecnologia social de sanção aflitiva ou premial orientadora de setores relevantes via atividade contratual, ordenadora, gerencial ou fomentadora.” (Iorio Aranha, 2019, p. 41)

A regulação pode ser compreendida também a partir dos métodos jurídico-regulatórios disponíveis ao regulador. Para compreender esses métodos, é importante realizar a distinção teórica entre autonomia e heteronomia.

A autonomia — do grego αὐτο-, "de si mesmo" e νόμος, "lei" — remete, conforme Kant, à faculdade racional humana “de atuar de acordo com leis que o agente dá a si mesmo, mediante as quais ele age independentemente de ser determinado por causas estranhas.” (Ramos, 2008)

A heteronomia — do grego ἕτερος-, "outro", “diferente” e νόμος, "lei" — também segundo Kant, “é possível quando, na ausência de uma lei que o sujeito dá a si mesmo, a pessoa

por passividade, covardia ou violência externa submete-se à lei e ao juízo de outrem, renunciando ao uso autônomo da razão em toda a sua capacidade e alcance.” (Ramos, 2008)

O conceito de autonomia está ligado à ideia de coação interna, e o conceito de heteronomia está ligado à ideia de coação externa. Materialmente, a regulação se dá por meio dessas duas técnicas principais.

Em termos jurídicos, a coerção externa se refere à atuação sancionadora da norma jurídica sobre os comportamentos desviantes do regulado, e está ligada à ideia de regulação de conformidade ou de comando e controle. “Para teorias apoiadas na percepção do direito como coerção extrínseca, ele — o direito — somente se realiza quando descumprido” (Iorio Aranha, 2019, p. 48).

Já a coerção interna corresponde à criação de uma rede de incentivos regulatórios que promove a cooperação e a confiança. É o que se chama de regulação por incentivos e está ligada à ideia de regulação aspiracional (Iorio Aranha, 2019).

É evidente que o uso de técnicas de coação interna — ou intrínseca — ou de coação externa — ou extrínseca — deve ser complementar. Não é possível realizar a regulação a partir de apenas um desses paradigmas.

Há algumas representações visuais interessantes que expressam a complementaridade entre esses métodos regulatórios.

Figura 1 — Exemplo de uma pirâmide de *enforcement*

Fonte: Ayres e Braithwaite *apud* Kolieb (2015)

A primeira é a pirâmide de *enforcement* de Ayres e Braithwaite. Essa pirâmide é uma representação visual e intuitiva das técnicas regulatórias de conformidade a serem utilizadas. Na base da pirâmide estão as metodologias mais dialógicas, inclusivas e colaborativas para garantir *compliance*, ou conformidade, com a lei (Braithwaite, 2002) À medida que se sobe na pirâmide, as intervenções tornam-se cada vez mais punitivas.

Figura 2 — Imagem da pirâmide de *enforcement* sob uma perspectiva de regulação responsiva, mostrando presunções feitas pelo regulador a respeito da entidade regulada

Fonte: Braithwaite *apud* Kolieb (2015)

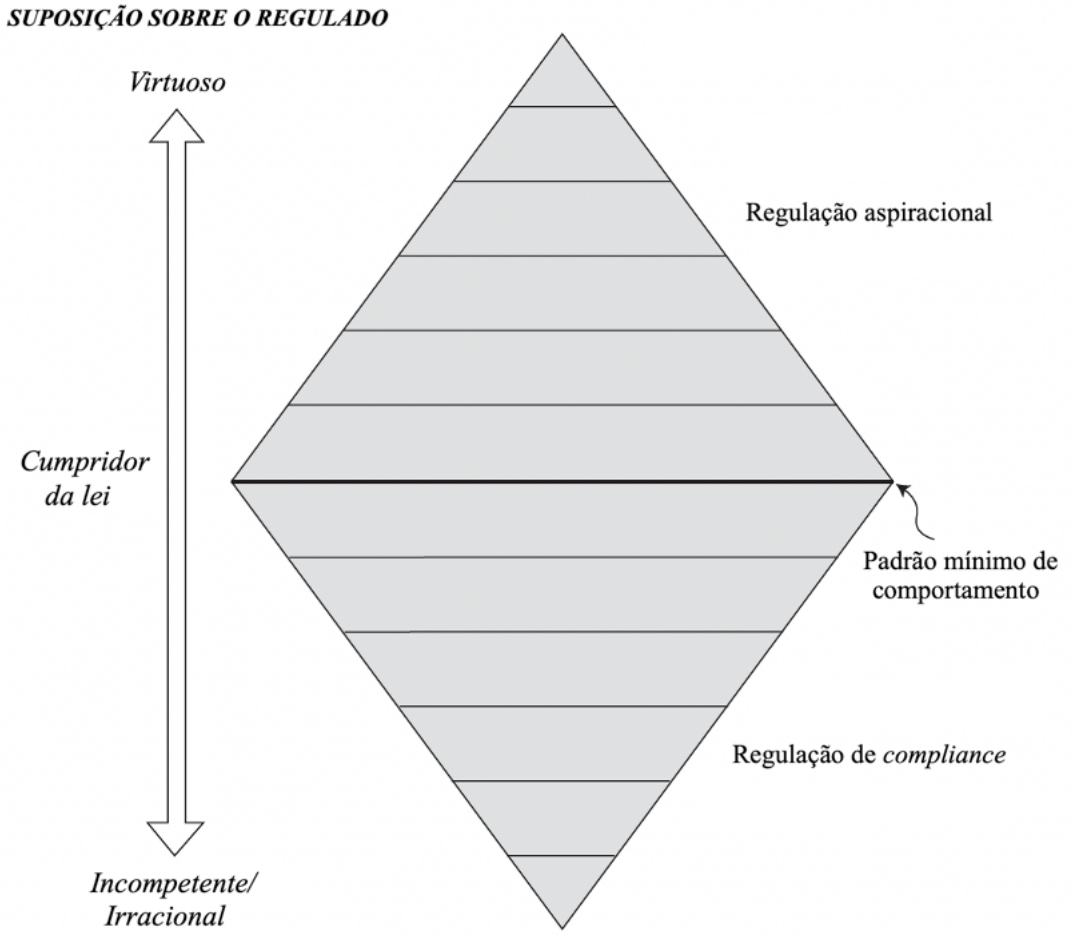
A segunda é a pirâmide de utilidade de Braithwaite. Essa pirâmide dispõe o tipo de resposta regulatória exigida em função de como o regulador enxerga o regulado.

Na base da pirâmide encontra-se a resposta "justiça restaurativa" para os atores que são virtuosos. A ideia da justiça restaurativa é, a partir do diálogo e de incentivos intrínsecos, convencer o regulado a adotar ações mais virtuosas.

No meio da pirâmide está a dissuasão. Esse método de regulação possibilita resultados positivos com atores racionais, ou seja, aqueles que entendem que precisam seguir a lei para continuarem no mercado.

No topo da pirâmide está a incapacitação. Esse tipo de ação está reservado àqueles atores incapazes ou irracionais, que são insensíveis ao constrangimento das normas jurídicas e buscam, de qualquer forma, burlar a lei.

Figura 3 — Diamante regulatório de Kolieb



Fonte: Kolieb (2015)

A terceira imagem é o Diamante Regulatório de Kolieb. Conforme se vê na figura acima, a partir de uma presunção do comportamento do regulado, o regulador apresenta diferentes respostas. Perceba-se que o comportamento do regulado é disposto num espectro que vai de incompetente ou irracional, num extremo, passando por cumpridor da lei e chegando em virtuoso, no extremo oposto.

Esse modelo destaca que o comportamento dos regulados que meramente seguem a lei por, racionalmente, temerem as sanções decorrentes do descumprimento, é considerado um padrão mínimo de comportamento — ou *minimum standards*, no idioma original.

Aos regulados incompetentes ou irracionais, a resposta do regulador está em sanções jurídicas progressivamente mais duras e onerosas. Esse tipo de resposta, no modelo, é chamado de regulação de *compliance*, que corresponde aos mecanismos para impor, juridicamente, o padrão mínimo de comportamento.

Aos regulados virtuosos, a resposta do regulador está em incentivos progressivamente mais interessantes e incentivadores à sua atuação no mercado. Esse tipo de resposta, no modelo, é chamado de regulação aspiracional, que corresponde a incentivos para que os regulados que buscam exceder os *minimum standards* de um dado espaço regulatório.

É extremamente importante que a regulação das tecnologias de inteligência artificial siga um modelo de regulação responsiva, com claros incentivos ao comportamento virtuoso, em vez de focar apenas nas punições pelo desvio da norma. Essa é uma recomendação importante para qualquer espaço de regulação, mas mais válido ainda para um espaço em que há grande potencial de inovação e constante mudança, pois é necessário criar uma cultura que incuta nos desenvolvedores e provedores de aplicações de inteligência artificial os princípios tendentes a um comportamento ético a cada nova descoberta, avanço ou produto. Aumentar o número de regulados virtuosos por meio de regulação inteligente economizará recursos e trará resultados ótimos ao setor. Caso isso não seja feito de forma paulatina, a probabilidade de embates entre reguladores e regulados a cada rodada de inovação é muito alto, com constantes judicializações e prejuízos aos usuários.

Um ponto central para a regulação são as normas que efetivamente fazem a regulação existir no mundo jurídico e que comunicam aos regulados as diretrizes elaboradas pelo regulador. Mesmo numa situação de sistemas de incentivos ou de normas não-estatais, construídas dialogicamente junto a determinado setor regulado, o papel das normas jurídicas é crucial.

Diferentes tipos de normas são inclusos no ordenamento jurídico de formas distintas. Conforme o paradigma proposto por Hans Kelsen (1967) em seu famoso modelo teórico piramidal, no topo do ordenamento jurídico há a *grundnorm*, que funda um dado ordenamento jurídico e da qual emanam todas as outras normas, em cascata, estando a constituição em um nível superior. Destaca-se que o objetivo deste trabalho não é discutir a natureza da *grundnorm*, mas apenas utilizar o modelo teórico de Kelsen para elucidar questões processuais e legislativas do ordenamento jurídico.

As normas de uma constituição gozam, de forma geral, de maior proteção contra mudanças legislativas, principalmente por meio da exigência de quóruns mais elevados e processos legislativos mais complexos e que exigem maior consenso político para sua alteração.

Abaixo da constituição há leis exaradas pelo corpo legislativo de um dado Estado, ou de determinado nível federativo. Diferentes leis podem ter diferentes tipos de requerimentos em termos de processo legislativo, a depender de aspectos materiais ou formais.

Por fim, abaixo das leis há as normas infralegais, que seguem processos de nomogênese mais simplificados, e muitas vezes realizados pelo Poder Executivo.

Considerando-se o que se sabe sobre a regulação — que deve responder de forma ágil às ações dos regulados — fica claro que não seria possível que toda regulação estatal estivesse inclusa nos níveis constitucional ou legal do ordenamento jurídico. De fato, "também se insere no rol de pressupostos do Estado Regulador o gerenciamento normativo da realidade regulada via administração das leis, para plena aplicação do princípio do *due process of law*, tão bem traduzido por Miguel Reale como a devida atualização do direito." (Iorio Aranha, 2019, p. 23) A regulação, portanto, inclui a atualização do direito de forma tempestiva, em especial em setores em constante inovação, como o de inteligência artificial. A probabilidade de que os reguladores consigam acompanhar os desenvolvimentos das tecnologias em tempo real é muito baixa, então deve-se, ao menos, aumentar a agilidade com que normas setoriais possam ser exaradas pelos corpos técnicos reguladores responsáveis de forma direta, garantido sempre o contato próximo aos regulados.

Quando se discute se um setor deve ou não ser regulado, normalmente se consideram fatores como a existência de monopólios *de facto* ou *de jure*; a existência de sérias barreiras de entrada no mercado; a existência de custos fixos pesados; ou outros fatores que limitam, materialmente, a concorrência mercadológica e que podem prejudicar a qualidade do fornecimento do serviço em questão (Bacache-Beauvallet; Perrot, 2017). Apesar de parecer existir um risco de concentração de mercado em agentes que consigam mobilizar mais infraestrutura para realização de cálculos computacionais para o "treino" das máquinas, à medida que ocorre a popularização da tecnologia de inteligência artificial e a disseminação da técnica, essas barreiras parecem se tornam menos preponderantes.

Além disso, há outras justificativas teóricas para a regulação de setores que sejam socialmente sensíveis. Diz Odete Medauar:

> "[...] regulação não visa exclusivamente à atividade econômica e aos serviços públicos. Podem ser objeto de regulação, como ocorre, por exemplo, na França, na Itália, na

> Inglaterra, os chamados setores sensíveis da vida social, como preservação de dados pessoais, segurança do trabalho, acesso a documentos, relações raciais. São relações e valores não econômicos, fugindo, portanto, à ideia de que regulação inclui necessariamente concorrência." (2002)

Pelos impactos jurídicos e riscos apresentados na seção anterior, é evidente que a inteligência artificial facilmente se enquadra no rol de setores sensíveis da vida que merecem atenção regulatória com base nos riscos específicos que os reguladores e os regulados compreendem que essas tecnologias oferecem.

Aliás, vale destacar que, em 2016, apenas uma lei referente à inteligência artificial aprovada no mundo. Em 2022 foram 37 (Stanford, [2023]).

Quase três a cada quatro americanos é favorável à regulação da IA (Orth, 2023). Ainda, como evidenciado pela carta aberta de líderes do ramo tecnológico, mencionada em seção anterior, existe uma espécie de consenso quanto à necessidade de regulação entre as grandes empresas capazes de efetivamente construir as tecnologias de inteligência artificial de ponta — ao menos quanto à sua comunicação institucional — e legisladores ao redor do mundo. Parece que a conjuntura, então, torna quase inevitável, politicamente, a regulação dessas tecnologias.

Uma via para adentrar a sensibilidade regulatória da temática da inteligência artificial e compreender as diferentes soluções regulatórias é estudar tanto propostas legislativas correntes no Brasil quanto propostas do Parlamento Europeu.

Escolheu-se a perspectiva do Brasil para este estudo de forma natural, já que é nossa jurisdição, e as legislações do Brasil impactam diretamente a prática jurídica e regulatória em nosso contexto e deve ser conhecida por todos os juristas brasileiros.

Já a escolha da proposta legislativa da União Europeia se justifica porque é a primeira tentativa no mundo de regulação da inteligência artificial; é, também, a mais relevante globalmente; é uma proposta que claramente inspirou os projetos de lei na mesma temática que hoje tramitam no Brasil; e a União Europeia já inspirou o Brasil no passado quanto à produção legislativa em temas relacionados à tecnologia, como é o caso das similaridades entre a Lei Geral de Proteção de Dados (Lei 13.709/2018) e a *General Data Protection Regulation*, promulgada pelo Parlamento Europeu.

O fio condutor da análise será a legislação proposta no âmbito do Parlamento Europeu, tendo em vista que é cronologicamente anterior, a quem o projeto de lei brasileiro será comparado.

4.1 AI ACT do Parlamento Europeu

O Artificial Intelligence Act, que atualmente tramita no Parlamento Europeu, é considerado a primeira tentativa no mundo de regulação ampla da inteligência artificial. (Parlamento Europeu, 2023).

Esta proposta legislativa é, por excelência, a primeira e mais significativa reação jurídica aos desenvolvimentos e desdobramentos das tecnologias baseadas em inteligência artificial, sendo proposta em 21 de abril de 2021. A proposta legislativa terá alguns de seus pontos mais candentes enumerados abaixo, já com as emendas aprovadas em 14 de junho de 2023, que mudaram substancialmente muitos aspectos da proposta. A intenção deste texto, por óbvio, não é realizar uma análise completa da extensa legislação — o que foge do escopo do trabalho — mas focar nas interações entre as reações jurídico-regulatórias e os impactos e riscos oferecidos pelas tecnologias de inteligência artificial.

A ideia central da proposta é de lidar com as diferentes aplicações da inteligência artificial conforme os diferentes riscos que cada aplicação pode oferecer para os usuários e para a sociedade como um todo. Há também aplicações que são proibidas, ou seja, cujo risco não é tolerável ou que são, em si, consideradas contrárias à vida em sociedade.

Há de se lembrar que, como se trata de processo legislativo, o rito inclui complexidades que interferem nas discussões e nos resultados da regulação. Por exemplo, o texto original proposto pela Comissão traz uma abordagem para a aceitabilidade de riscos que é distinta daquela adotada pelo Conselho e pelo Parlamento, como expresso em suas emendas respectivas propostas (Fraser; Villarino, 2023). De fato, as medidas de mitigação de risco preconizadas pela Comissão não pareciam levar em conta a dimensão de custos, e apenas preconizava que os riscos fossem mitigados "*as far as possible*", ou "o quanto possível". Já a posição do Conselho e do Parlamento incluem uma dimensão de razoabilidade, incluindo, assim, análises de custo-benefício (Fraser; Villarino, 2023).

Dito isso, abaixo expõe-se o estado atual do projeto, já considerando as emendas propostas.

4.1.1 Banimento de aplicações com risco de serem utilizadas para desinformação e manipulação

Entre as aplicações proibidas, o primeiro grupo refere-se àquelas que utilizam técnicas subliminares — isto é, além da consciência da pessoa —, manipulativas ou enganadoras para distorcer seu comportamento de forma a provavelmente causar dano físico ou psicológico a ela ou a outros. Em especial, foi destacada a proibição das aplicações que visem explorar vulnerabilidades de grupos específicos, incluindo por conta de deficiência física ou mental. Esse risco abordado pela proposta legislativa europeia pode ser categorizado como risco de manipulação. Quanto à tipologia preconizada por este trabalho, seria um risco a direitos fundamentais.

Apesar de essa ser, textualmente, uma proibição bem razoável, com certeza não se aplica apenas à inteligência artificial; de fato, é aplicável para além do campo da tecnologia, sendo uma diretriz lógica para qualquer tipo de comunicação de massa. A problemática, em si, entra na avaliação do que efetivamente configurará, na prática, uma técnica subliminar ou manipulação. Isso gera um risco político no sentido de ser possível enquadrar discursos de oposição política como manipulativos.

Aliás, em falar de manipulação, o estudo ético do tema traz diversas perspectivas do que efetivamente é manipulação (Noggle, 2023). Uma das definições possíveis de manipulação por técnica subliminar é pensar em algo que consiga furtivamente causar convencimento em outra pessoa sem verdadeira deliberação racional (Noggle, 2023). Vê-se, então, que há espaço significativo para interpretação do que é ou não influenciar outra pessoa subliminarmente, e há mais espaço ainda para definir o que se considera, de forma geral, manipulação.

4.1.2 Banimento de aplicações com risco de serem utilizadas para "social scoring" e sistemas de identificação biométrica para o combate ao crime

Outra das aplicações proibidas é a utilização de dados colhidos para a avaliação ou classificação de pessoas naturais para pontuá-las a partir de uma metodologia de "*social score*" que possa levar ao tratamento desfavorável de algumas pessoas em relação a outras, em especial em contextos sociais diversos daqueles em que os dados foram colhidos ou de forma

injustificada ou desproporcional. Aqui, o risco também pode ser categorizado no rol daqueles ligados à violação de direitos fundamentais.

De fato, a prática de pontuação social, em especial realizada por meios automatizados de vigilância massiva, é um grande perigo para a liberdade humana, então sua proibição não apenas por parte de autoridades públicas, mas também de particulares, é bastante razoável.

Por fim, o estado atual da proposta legislativa realizou um banimento *tout court* de sistemas de identificação biométrica. A proposta original continha exceções para a utilização desse tipo de identificação para a proteção de vítimas de crimes, incluindo crianças desaparecidas, para a prevenção de terrorismo e para a persecução penal, incluindo detecção e localização de infratores ou suspeitos de cometimento de crimes, o que foi completamente removido com as emendas mais atuais.

Não se trata de uma questão óbvia ou simples, pois os sistemas de identificação biométrica em massa podem ser utilizados complementarmente aos sistemas de segurança vigentes, de forma virtuosa, ou podem também ser um patamar anterior à implementação de um sistema de pontuação social, o que é sempre indesejável. Entretanto, o banimento generalizado e sem exceções é no mínimo preocupante e mostra que falta flexibilidade e entendimento para resolver a questão. Parece ser mais razoável existir uma entidade pública e transparente, acessível por todos os interessados, que regule e observe de muito perto sistemas de segurança para as exceções originalmente propostas.

4.1.3 Aplicações de alto risco, mas não banidas

A proposta legislativa prevê um rol de aplicações de alto risco, isto é, que sofrerão mais intensa regulação, mas que não são *a priori* banidas ou proibidas.

A lista de aplicações de alto risco foi determinadas a partir de fatores previstos na própria legislação, em especial o alcance dos efeitos produzidos pela aplicação em questão, a reversibilidade desses efeitos e os riscos materiais de dano ou impacto adverso.

O primeiro grupo de aplicações designadas como de alto risco pela proposta legislativa são aquelas que se baseiam no uso de biometria, mas que não envolvam os casos tendentes a "social scoring".

Essa aplicação representa um alto risco de ferir os direitos de já que o uso indevido de dados pessoais por governos ou corporações é preocupante. Por fim, com a ubiquidade de

câmeras e outros aparatos biométricos, os usuários do sistema poderão ter sua privacidade violada no sentido de que seus dados são coletados sem seu consentimento. Há também a possibilidade de violação da segurança dos dados por invasores. Esses riscos podem ser atribuídos à categoria daqueles relacionados a direitos fundamentais — pois lidam com a questão da privacidade — e à operação não pretendida.

Na mesma esteira, existe um risco nesse tipo de aplicação de reforço de vieses e de discriminação em referência a determinados grupos demográficos, e, devido à opacidade inerente aos sistemas de inteligência artificial, pode ser difícil identificar a ocorrência de decisões viesadas de antemão. O tipo de risco principal encontrado nessa situação é risco de compreensão.

O segundo grupo de aplicações de alto risco são aquelas usadas como componentes de segurança e gerenciamento do tráfego rodoviário e para o suprimento de água, gás, aquecimento, eletricidade e infraestrutura digital crítica.

O fator mais preponderante para a inclusão desse grupo são os impactos possíveis; na tipologia utilizada neste texto, esse risco está coadunado àquele referente à disrupção de sistemas essenciais à vida em sociedade — danos nesse tipo de infraestrutura, bem como sua má utilização, pode colocar em risco a vida de multidões e comprometer a estabilidade da sociedade. Essas aplicações apresentam riscos principalmente relacionados à operação não pretendida e a direitos. A operação não pretendida nesse caso pode se dar por ataques intencionais ou então ser uma derivação do risco de compreensão.

O terceiro grupo de aplicações de alto risco se refere àquelas usadas para fins de avaliação educacional ou vocacional. Esse tipo de avaliação lida com aspectos extremamente sensíveis, como a formação de um ser humano e seu direcionamento no mercado de trabalho. Um risco pernicioso seria de vieses que tenderiam a avaliar pessoas com determinadas características de forma mais positiva ou mais negativa.

O quarto grupo de aplicações de alto risco se refere àquelas que são usadas para o acesso ao emprego, ao recrutamento e ao desligamento de pessoas de seus empregos, e o quinto grupo de aplicações de alto risco se refere àquelas que podem ser utilizadas para definir ou restringir acesso a serviços públicos ou privados essenciais, ou a benefícios, como de serviços de emergência, ou pontuação de crédito.

O sexto grupo de aplicações de alto risco se refere àquelas que sejam utilizadas na aplicação e execução da lei (“*law enforcement”*). O sétimo grupo de aplicações de alto risco se

refere àquelas que sejam utilizadas em situações de migração, asilo e controle de fronteiras. O oitavo grupo de aplicações de alto risco se refere àquelas que sejam utilizadas na administração da justiça, ou seja, aplicações que ofereçam suporte para a autoridade judicial ou corpo administrativo em pesquisar e interpretar fatos e a lei, bem como em aplicar a lei a uma situação concreta.

Quanto aos grupos 3, 4, 5, 6, 7 e 8, observam-se três tipos de riscos: por ter potencial de ferir a equidade, podem ser categorizados como riscos a direitos fundamentais — as situações 6 e 8 foram inclusive diretamente abordadas em seção anterior; bem como apresentam riscos de operação não pretendida pelos desenvolvedores, bem como riscos referentes à compreensão quanto à operação da IA em si — quando não se conseguir explicar, por exemplo, a razão pela qual uma escolha foi feita em detrimento de outra.

O nono grupo de aplicações de alto risco se refere àquelas que sejam construídas para influenciar o resultado de uma eleição ou referendo. Neste caso há um risco de manipulação, que corresponde, na tipologia, àqueles que ferem direitos fundamentais.

O décimo grupo de aplicações de alto risco se refere àquelas que sejam utilizados por plataformas de redes sociais consideradas como muito grandes, conforme regulação pertinente. Como essa questão envolve pontos de responsabilização, vieses, opacidade, poder de corporações e privacidade, todas as quatro dimensões estão em jogo.

4.1.4 Obrigações gerais quanto à transparência

Algumas aplicações de inteligência artificial, independentemente de seu grau de risco, têm algumas obrigações quanto à transparência. Toda aplicação que interaja com um ser humano deve ser desenvolvida de forma que o ser humano seja avisado claramente que se trata de uma interação com inteligência artificial e se há suporte de um ser humano no gerenciamento do processo e quem é o responsável pelo processo decisório. Além disso, qualquer utilização permitida pela legislação de dados biométricos deve ser expressamente autorizada pelo usuário.

Outro ponto interessante é a disposição a respeito do que se chama *deep fake* — imagens, áudios e vídeos preparados digitalmente que mostram pessoas fazendo ou dizendo coisas que elas não fizeram. O uso de *deep fake* por qualquer aplicação de inteligência artificial deve ser notificado ao usuário de forma clara, tempestiva e visível.

4.1.5 Obrigações específicas para aplicações de alto risco

A proposta legislativa estabelece uma regulação de conformidade extensiva que as aplicações de alto risco devem seguir. Resume-se a seguir os principais pontos.

Em primeiro lugar, exige-se a implementação e documentação de um sistema de gerenciamento de riscos, que deve ser sistematicamente atualizado. A ideia é a mitigação e eliminação paulatina de riscos assim que identificados, quando possível, ou manejados e controlados. Nesse sentido, busca-se, pela documentação, além da mitigação global de riscos de diversos tipos, a redução de riscos relacionados à responsabilização: ao se conhecer de antemão os riscos específicos da operação e as condutas para lidar com eles, pode-se atribuir responsabilidade de forma mais racional em caso de realização do dano.

Outro ponto de importância é a implementação de práticas de gerenciamento e governança apropriados para treino, validação e teste de dados. Exige-se o exame de vieses que ofereçam risco para a saúde, segurança e direitos fundamentais, destacando-se aqueles vieses tendentes à discriminação de pessoas.

As aplicações de alto risco deverão promover a transparência, com instruções e documentação clara, atualizada e acessível ao usuário. Deverão também ser desenvolvidas tendo em vista interfaces com o ser humano de forma a possibilitar o gerenciamento das aplicações por parte de seres humanos, a compreensão do processo e até mesmo sua interrupção, se assim for necessário.

Conforme o artigo 43 da proposta, a avaliação de conformidade é um procedimento que pode ser realizado internamente, pelo próprio provedor a aplicação de IA de alto risco, ou por um terceiro autorizado. O regulador não busca, portanto, realizar uma intervenção que observe pessoalmente cada passo do regulado — que seria, por si só, impossível, considerando a quantidade de agentes e o ritmo da inovação de mercado — mas trata a todos, de início, como agentes que buscam cumprir a lei. A avaliação de conformidade é necessária para todos os sistemas de inteligência artificial categorizados como de alto risco, e deverá ser repetida toda vez que o sistema sofrer mudanças significativas.

A falha em cumprir com os requisitos da legislação, a depender da gravidade, pode chegar a uma punição por multa de 30 milhões de euros, ou 6% das receitas anuais globais para o ano financeiro precedente, o que for maior.

Do ponto de vista regulatório, a proposta cria a figura de uma "*sandbox* regulatória", que é um ambiente controlado de testes, experimentação e inovação para que as empresas mais facilmente consigam coadunar suas práticas aos requisitos da legislação. Esse instituto parece entrar na esteira das iniciativas regulatórias que visam a mais rápida atualização da regulação por meio do trabalho próximo aos regulados.

Criou-se uma estrutura de governança e conformidade que, a princípio, lança mão de *enforcement* em caso de descumprimento ou desrespeito às regras estabelecidas, mas que, de forma equilibrada, dá espaço para que o agente demonstre que tem interesse em cumprir a lei, o que parece se coadunar à pirâmide de *enforcement* de Ayres e Braithwaite. Uma forma de aprimoramento seria ter mais elementos que incluíssem o conceito de justiça restaurativa de Braithwarte ou mesmo uma dinâmica como a expressa no diamante regulatório de Kolieb, de forma a incentivar e estabelecer foco no desenvolvimento de comportamento virtuoso.

4.2 Projeto de Lei 2338/2023

O Projeto de Lei 2338/2023 (Senado Federal, 2023) é de uma linhagem de propostas legislativas que buscam lidar com o tema da inteligência artificial no Brasil. É a mais recente e mais completa, e por isso foi escolhida para constar deste trabalho. É também a mais influenciada pelo AI ACT do Parlamento Europeu.

O PL brasileiro organiza-se de forma muito similar à proposta legislativa original de 2021 do Parlamento Europeu, com diversos dispositivos e institutos que foram transplantados de forma direta. Apesar da boa intenção — de trazer as discussões mais avançadas sobre aplicações de inteligência artificial para o Brasil —, melhor seria a escrita de um texto original para esses pontos — se bem que há novidades interessantes sobre as quais discorre-se abaixo.

O Projeto de Lei tem um foco louvável nos direitos dos usuários. Garante, no art. 5º, direitos às pessoas afetadas por sistemas de inteligência artificial: o direito à informação prévia quanto à interação do usuário com a aplicação; o direito de explicação sobre decisão tomada pela aplicação; direito de contestação de decisões ou previsões que produzam efeitos jurídicos ou que impactam o interesse do afetado; direito à determinação e à participação humana nos processos decisórios das aplicações; direito à não discriminação e à correção de vieses discriminatórios; e direito à privacidade.

Note-se que não se usa, neste ponto, a palavra “usuário”, mas “pessoa afetada”, o que amplia significativamente o escopo do dispositivo e garante proteções relevantes não apenas para quem utiliza o sistema, mas para todos os afetados por sua operação.

A proposta de normativo brasileira organiza também a regulação a partir dos riscos que as aplicações de inteligência artificial podem oferecer.

O primeiro tipo de risco mencionado é o risco excessivo, e os sistemas categorizados dessa forma são completamente vedados. A lista daquelas aplicações que são tachadas de risco excessivo corresponde quase que exatamente à lista de aplicações proibidas no EU AI Act, mas no formato original de 2021. Isso significa que a proposta brasileira não inclui a vedação ampla contra manipulação, mas apenas contra técnicas de convencimento subliminar, e não inclui o banimento generalizado do uso de sistemas de identificação biométrica para propósitos de segurança pública.

Os sistemas de alto risco são categorizados, assim como na legislação europeia, por sua finalidade. A legislação brasileira traz todos aqueles mencionados no EU AI Act, inclusive os que foram inseridos por meio das emendas de junho de 2023, com diferenças muito limitadas. Até mesmo as hipóteses de identificação de novas aplicações de alto risco são muito próximas daquelas presentes na proposta europeia.

Quanto à conformidade regulatória, a proposta brasileira contém duas etapas principais: uma avaliação preliminar, obrigatória para todos os agentes que proveem qualquer sistema baseado em tecnologias de inteligência artificial, em que se avalia a categorização de risco; e a avaliação de impacto algorítmico, obrigatória apenas para os sistemas considerados de alto risco, que corresponde à “avaliação de conformidade” europeia, inclusive com a possibilidade de realização da avaliação por terceiros autorizados.

Uma novidade que provavelmente gerará discussão é quanto à responsabilidade civil. O Projeto de Lei prevê responsabilização civil objetiva por danos causados por sistemas de inteligência artificial de risco excessivo ou de alto risco. Quando o sistema de inteligência artificial não for de risco excessivo ou de alto risco, a culpa do agente causador será presumida e o ônus da prova será invertido em favor da vítima. Essa é a forma com que o PL trata, frontalmente, dos riscos relacionados à responsabilização.

Do ponto de vista regulatório, o PL brasileiro se inspirou na ideia de *sandbox* regulatório, bem como nos procedimentos em caso de falha em cumprir com os requisitos da legislação — a punição por desconformidade pode chegar a multa de 50 milhões de reais, ou

2% das receitas anuais no Brasil no último exercício financeiro, bem como restrição de acesso futuro a *sandbox* regulatório.

Assim como a proposta europeia, seria interessante que a regulação contivesse mais elementos que premiasse a boa conduta daqueles agentes que demonstram virtude no mercado, mas, de início, já contém elementos importantes para o início de uma trajetória de regulação responsiva.

5 DISCUSSÃO

Os impactos das tecnologias baseadas em inteligência artificial sobre a vida humana são multifacetados, com uma diversidade enorme de consequências em vários campos. Em toda a sociedade — e no âmbito do Direito, em particular — tanto pela infinidade de implicações do uso dessas tecnologias, quanto pelas aplicações instrumentais da inteligência artificial, é esperada uma transformação de escala monumental.

Tudo indica que profissionais de todas as indústrias e de todos os ramos — incluindo os operadores do Direito — utilizarão as tecnologias baseadas em inteligência artificial ao ponto de tornarem-se dependentes delas para maximização de sua produtividade, como hoje o são do computador, do e-mail e da internet. A substituição do ser humano por completo, entretanto, não é possível e nem desejável.

A partir dos riscos mapeados na literatura, foram propostas quatro categorias: (1) riscos de responsabilização: são riscos relacionados à dificuldade de responsabilização por erros ou danos cometidos por aplicações de IA; (2) riscos de operação não pretendida: são riscos relacionados à operação não pretendida pelos desenvolvedores da aplicação de IA; (3) riscos de compreensão: são riscos relacionados à falta de compreensão dos desenvolvedores ou dos usuários quanto à operação da IA; e (4) riscos a direitos fundamentais: são riscos relacionados a ameaças a direitos fundamentais, como vida, incolumidade física, privacidade, patrimônio e trabalho. Essas categorias abrangem os riscos conhecidos e os organiza por afinidade temática, o que facilita a análise de propostas legislativas ou análises de risco em geral.

A forma mais adequada de compreender a tipologia proposta é no sentido de aplicação de diferentes níveis de análise para um mesmo fenômeno, e não como tipos concorrentes e mutuamente excludentes. Todas as quatro dimensões devem sempre ser empregadas para a compreensão integral do fenômeno, mas é comum que uma ou outra sobressaia-se, a depender da situação.

De fato, os riscos que foram identificados nesta monografia foram, ao menos, endereçados pelas propostas legislativas analisadas. Aqueles riscos identificados na literatura como percebidos pelos entrevistados como de maior severidade foram todos nominalmente abordados na legislação analisada — isto é, riscos quanto ao tratamento não equitativo, riscos de viés e riscos referentes à proteção/privacidade de dados.

Pode-se dizer que tanto os provedores de aplicações de inteligência artificial quanto os reguladores governamentais admitem não poder prever os futuros desenvolvimentos e avanços das tecnologias de inteligência artificial — o que se reflete nas reações regulatórias mais relevantes da atualidade, que parecem ser em geral flexíveis, com pontos específicos para promover e proteger a inovação. Isso é observado tanto na proposta europeia quanto na proposta brasileira.

Em pontos específicos, como no caso da utilização de tecnologias biométricas de inteligência artificial no contexto da segurança pública, o debate apresenta-se acirrado, e é possível que os encaminhamentos finais das legislações apontem para resultados regulatórios mais restritivos.

Os impactos dessas tecnologias serão modulados, portanto, pela regulação que já desponta no horizonte. Independentemente de que tecnologias sejam criadas, a sociedade e os reguladores precisarão entender que o propósito da inteligência artificial, afinal, é o de facilitar a vida humana, e não de substitui-la.

REFERÊNCIAS

ALLYN, B. **A robot was scheduled to argue in court, then came the jail threats**. NPR. 25 jan. 2023. Disponível em: https://www.npr.org/2023/01/25/1151435033/a-robot-was-scheduled-to-argue-in-court-then-came-the-jail-threats. Acesso em: 20 set. 2023.

ATKINSON, R. D. **"It's Going to Kill Us!" and Other Myths About the Future of Artificial Intelligence**. ITIF. 6 jun. 2016. Disponível em: https://itif.org/publications/2016/06/06/its-going-kill-us-and-other-myths-about-future-artificial-intelligence/. Acesso em: 20 set. 2023.

BACACHE-BEAUVALLET, M.; PERROT, A. **Economic Regulation: Which Sectors to Regulate and How?** Cairn International Edition. 2017. Disponível em: https://www.cairn-int.info/article-E_NCAE_044_0001--economic-regulation-which-sectors-to.htm. Acesso em: 20 set. 2023.

BENDER, M. E.; GEBRU; T.; MCMILLAN-MAJOR, A.; SHMITCHEL, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? **FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency**. Nova York, NY: 2021.

BERNERS-LEE, T.; CAILLIAU, R.; GROFF, J. F.; POLLERMANN, B. World-Wide Web: the information universe. **Internet Research**, 2(1), 52-58. 1992.

BIRHANE, B. **The unseen Black faces of AI algorithms**. Nature. 2022. Disponível em: https://www.nature.com/articles/d41586-022-03050-7. Acesso em: 11 dez. 2023.

BLOOMBERG LAW. **Write a Better Legal Brief in Less Time**. Bloomberg Law. 11 ago. 2023. Disponível em: https://pro.bloomberglaw.com/brief/how-to-write-a-legal-brief/. Acesso em: 20 set. 2023.

BRAGA, A.; LOGAN, R. The Emperor of Strong AI Has No Clothes: Limits to Artificial Intelligence. **Information**, v. 8, n. 4, p. 156. 2017. Disponível em: https://www.mdpi.com/2078-2489/8/4/156. Acesso em: 20 set. 2023.

BRAITHWAITE, J. **Restorative Justice and Responsive Regulation**. Oxford: Oxford University Press, 2002.

BRASIL. **Constituição da República Federativa do Brasil.** Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Acesso em: 20 set. 2023.

_____. **Lei nº 11.419, de 19 de dezembro de 2006**. Dispõe sobre a informatização do processo judicial; altera a Lei nº 5.869, de 11 de janeiro de 1973 — Código de Processo Civil; e dá outras providências. Brasília, DF: Presidência da República, 2006. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/l11419.htm. Acesso em: 20 set. 2023.

BROENS, M. C.; MORAES, J. A.; CORDERO, A. F. Technology and society: The impacts of Internet of Things on individual's daily life. **Cognitive Science: Recent advances and recurring problems**. Wilmington, Delaware: Vernon Press, 2017.

BROOKS, R. **Artificial Intelligence is a tool, not a threat**. Rethink Robotics. 2014. Disponível em: https://web.archive.org/web/20141112130954/http:/www.rethinkrobotics.com/artificial-intelligence-tool-threat/. Acesso em: 23 set. 2023.

BROWER, T. **People Fear Being Replaced By AI And ChatGPT**: 3 Ways To Lead Well Amidst Anxiety. Forbes. 5 mar. 2023. Disponível em: https://www.forbes.com/sites/tracybrower/2023/03/05/people-fear-being-replaced-by-ai-and-chatgpt-3-ways-to-lead-well-amidst-anxiety/?sh=242a310e7fe6. Acesso em: 20 set. 2023.

BURRELL, J. How the machine “thinks”: Understanding opacity in machine learning algorithms. **Big Data & Society**, v. 3, n. 1, p. 1—12. 5 jan. 2016. Disponível em: https://journals.sagepub.com/doi/full/10.1177/2053951715622512. Acesso em: 20 set. 2023.

BUTTERICK, M. **Stable Diffusion Litigation**. Stable Diffusion Litigation. 13 jan. 2023. Disponível em: https://stablediffusionlitigation.com/. Acesso em: 20 set. 2023.

CASTRO, C. What’s Wrong with Machine Bias. **Ergo, an Open Access Journal of Philosophy**, v. 6, n. 20191108. 11 jul. 2019. Disponível em: https://quod.lib.umich.edu/e/ergo/12405314.0006.015/--what-s-wrong-with-machine-bias?rgn=main;view=fulltext. Acesso em: 20 set. 2023.

CLIO. **Legal Trends Report 2021**. Clio. 2021. Disponível em: https://www.clio.com/resources/legal-trends/. Acesso em: 20 set. 2023.

CONSELHO NACIONAL DE JUSTIÇA. **Há déficit de 19,8% de juízes no Brasil.** CNJ. 14 set. 2017. Disponível em: https://www.cnj.jus.br/ha-deficit-de-19-8-de-juizes-no-brasil/. Acesso em: 20 set. 2023.

_____. **Resolução n° 325, de 28 de junho de 2020**. CNJ. 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3365. Acesso em: 20 set. 2023

COSMO, L. **Google Engineer Claims AI Chatbot Is Sentient: Why That Matters**. Scientific American. 2022. Disponível em: https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/. Acesso em: 20 set. 2023.

COZMAN, F. G.; KAUFMAN, D. Viés no aprendizado de máquina em sistemas de inteligência artificial: a diversidade de origens e os caminhos de mitigação. **Revista USP**, n. 135, p. 195—210. 22 dez. 2022. Disponível em: https://www.revistas.usp.br/revusp/article/view/206235/189877. Acesso em: 20 set. 2023.

CREPALDI, T.; GOES, S. **Justiça brasileira alcança marca de 80 milhões de processos em tramitação**. Conjur. 30 jun. 2022. Disponível em: https://www.conjur.com.br/2022-jun-30/poder-decide-faz. Acesso em: 20 set. 2023.

DIETVORST, B. J.; SIMMONS, J. P.; MASSEY, C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. **Journal of Experimental Psychology**, General, 144(1), 114. Washington, DC: American Psychological Association, 2015.

FETZER, J. H. **What is Artificial Intelligence?** pp. 3-27. Dordrecht: Springer Netherlands, 1990.

FLOWERS, J. C. Strong and Weak AI: Deweyan Considerations. **AAAI spring symposium: Towards conscious AI systems**, vol. 2287, No. 7. Washington, DC: Association for the Advancement of Artificial Intelligence, 2019.

FRANKENFIELD, J. **What is a hash? Hash functions and cryptocurrency mining.** Investopedia. 2023. Disponível em: https://www.investopedia.com/terms/h/hash.asp. Acesso em: 21 nov. 2023.

FRASER, H.; VILLARINO, J. M. B. Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough. **European Journal of Risk Regulation**, 1-16. Cambridge: Cambridge University Press, 2023.

FUNDAÇÃO GETULIO VARGAS. **Estudo sobre o Judiciário Brasileiro.** FGV. 2019. Disponível em: https://ciapj.fgv.br/sites/ciapj.fgv.br/files/estudo_da_imagem_do_judiciario_brasileiro.pdf. Acesso em: 20 set. 2023.

FUTURE OF LIFE INSTITUTE. **Pause Giant AI Experiments: An Open Letter**. Future of Life Institute. 2023. Disponível em: https://futureoflife.org/open-letter/pause-giant-ai-experiments/. Acesso em: 20 set. 2023.

GEIST, E. M. It's already too late to stop the AI arms race — We must manage it instead. **Bulletin of the Atomic Scientists**, 72(5), 318-321. Chicago, IL: Bulletin of the Atomic Scientists, 2016.

GOODRICH, J. **How IBM's Deep Blue Beat World Champion Chess Player Garry Kasparov**. IEEE. 25 jan. 2021. Disponível em: https://spectrum.ieee.org/how-ibms-deep-blue-beat-world-champion-chess-player-garry-kasparov. Acesso em: 20 set. 2023.

GRANULO, A.; FUCHS, C.; PUNTONI, S. Preference for human (vs. robotic) labor is stronger in symbolic consumption contexts. **Journal of Consumer Psychology**, 31(1), 72-80. Nova York, NY: John Wiley & Sons, 2021.

HAVICK, J. The impact of the Internet on a television-based society. **Technology in Society,** 22(2), 273-287. Amsterdam, Países Baixos: Elsevier BV, 2000.

HINTON, G. E. How neural networks learn from experience. **Scientific American**, 267(3), 144-151. Nova York, NY: 1992.

HUTSON, M. The opacity of artificial intelligence makes it hard to tell when decision-making is biased. **IEEE Spectrum**, v. 58, n. 2, p. 40-45. Fev. 2021. Disponível em: https://ieeexplore.ieee.org/abstract/document/9340114. Acesso em: 20 set. 2023.

IORIO ARANHA, M. **Manual de Direito Regulatório**. Fundamentos de Direito Regulatório. Londres: Laccademia Publishing, 2019.

KANG, C. **OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing**. The New York Times. 2023. Disponível em: https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html. Acesso em: 20 set. 2023.

KELSEN, H. **Pure theory of law**. Berkeley: University of California Press, 1967.

KOLIEB, J. **When To Punish, When To Persuade And When To Reward**: Strengthening Responsive Regulation With The Regulatory Diamond. 2015. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2698498. Acesso em: 20 set. 2023.

LAGE, F. C. **Manual da Inteligência Artificial no Direito Brasileiro**. 2ª ed. Salvador, BA: Editora Juspodivm, 2022.

MALIHA, G; PARIKH, R. **Who is liable when AI kills?** Scientific American. 20 jun. 2022. Disponível em: https://www.scientificamerican.com/article/who-is-liable-when-ai-kills/. Acesso em: 20 set. 2023.

MEDAUAR, O. Regulação e autorregulação. **Revista de direito administrativo,** 228. Rio de Janeiro: FGV, 2002. Disponível em: https://bibliotecadigital.fgv.br/ojs/index.php/rda/article/download/46658/44479. Acesso em: 20 set. 2023.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.

MORIKAWA, M. Firms' expectations about the impact of AI and robotics: evidence from a surgey. **Economic Inquiry**, v. 55, n. 2, p. 1054-1063. Hoboken, NJ: Wiley-Blackwell, 2016.

NOGGLE, R. **The Ethics of Manipulation**. Stanford Encyclopedia of Philosophy Archive. 2020. Disponível em: https://plato.stanford.edu/archives/sum2020/entries/ethics-manipulation/. Acesso em: 20 set. 2023.

OARD, D.W.; BARON, J.R.; HEDIN, B.; LEWIS, D. D.; TOMLINSON S. Evaluation of information retrieval for e-discovery. **Artif Intell Law**. Norwell, MA: Springer, 2010.

ORTH, T. **Americans are divided on AI's societal impact, but most support government regulation**. YouGov. 2023. Disponível em: https://today.yougov.com/politics/articles/45747-americans-are-divided-artificial-intelligence-poll?redirect_from=%2Ftopics%2Fpolitics%2Farticles-

reports%2F2023%2F05%2F25%2Famericans-are-divided-artificial-intelligence-poll. Acesso em: 20 set. 2023.

PAGE, J.; BRAIN, M.; MUKHLISH, F. The Risks of Low Level Narrow Artificial Intelligence. **2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR).** Shenyang, China: 2018.

PARLAMENTO EUROPEU. **EU AI Act**. 2023. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html. Acesso em: 20 set. 2023.

PETROPOULOS, G. **The dark side of artificial intelligence: manipulation of human behaviour**. Bruegel. 2022. Disponível em: https://www.bruegel.org/blog-post/dark-side-artificial-intelligence-manipulation-human-behaviour. Acesso em: 20 set. 2023.

PINHEIRO, P. P. G. Contratos digitais ou eletrônicos: apenas um meio ou uma nova modalidade contratual? **Revista dos Tribunais**, v. 966, p. 21-40. São Paulo: 2016.

RAMOS, C. A. Coação e autonomia em Kant: as duas faces da faculdade de volição. **ethic@ - An international Journal for Moral Philosophy**, v. 7, n. 1. 16 dez. 2008. Disponível em: https://periodicos.ufsc.br/index.php/ethic/article/view/1677-2954.2008v7n1p45/16080. Acesso em: 20 set. 2023.

RIBEIRO, L.; MENDIZABAL, O. **Introdução à Blockchain e Contratos Inteligentes**: Apostila para Iniciante. Florianópolis: UFSC/INE, 2021. Disponível em: https://repositorio.ufsc.br/handle/123456789/221495. Acesso em: 20 set. 2023.

ROUSE, M. **Machine Bias.** Techopedia. 12 out. 2023. Disponível em: https://www.techopedia.com/definition/31621/weak-artificial-intelligence-weak-ai. Acesso em: 6 nov. 2023.

SALGADO, J. C. Os direitos fundamentais. **Revista Brasileira Estudos Políticos**, 82, 15. Belo Horizonte, MG: UFGM, 1996.

SCHUMPETER, J. A. **Capitalism, Socialism and Democracy**. Nova York: Harper & Brothers, 1942.

SEARLE, J. R. Minds, brains, and programs. **Behavioral and brain sciences.** Cambridge, Inglaterra: Cambridge University Press, 1980.

SENADO FEDERAL. **Projeto de Lei 2338/2023** — Dispõe sobre o uso da Inteligência Artificial. 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 20 set. 2023.

SLOMAN, A. The irrelevance of Turing machines to artificial intelligence. **Computationalism: new directions**. p. 87-127. Cambridge, MA: MIT Press, 2002.

STANFORD. **The AI Index**. Stanford. [2023]. Disponível em: https://aiindex.stanford.edu/. Acesso em: 20 set. 2023.

STUART, R.; PETER, N. 26.3: The Ethics and Risks of Developing Artificial Intelligence. **Artificial Intelligence: A Modern Approach**. Hoboken, NJ: Prentice Hall, 2009. Disponível em: https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf. Acesso em: 11 dez. 2023.

SUPERIOR TRIBUNAL DE JUSTIÇA. **A era digital**. Brasília: STJ, [20--]. Disponível em: https://www.stj.jus.br/sites/portalp/Institucional/Historia/A-era-digital. Acesso em: 2 ago. 2023.

TAN, O. **How Does A Machine Learn?** Forbes. 2 mai. 2017. Disponível em: https://www.forbes.com/sites/forbestechcouncil/2017/05/02/how-does-a-machine-learn/?sh=4481a1d37441. Acesso em: 20 set. 2023.

TEIXEIRA, S., RODRIGUES, J., VELOSO, B., & GAMA, J. An Exploratory Diagnosis of Artificial Intelligence Risks for a Responsible Governance. **Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance,** pp. 25-31. Guimarães, Portugal: United Nations University, 2022.

THOMSON, J. J. The Trolley Problem. **The Yale Law Journal**, v. 94, n. 6, p. 1395-1415. New Haven, CT: The Yale Law Journal Company, Inc, 1985.

VINCENT, J. **AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit.** The Verge. 16 jan. 2023. Disponível em: https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart. Acesso em: 20 set. 2023.

WALKER, J. **On Legal AI**: Um rápido tratado sobre a Inteligência Artificial no Direito. São Paulo, SP: Thomson Reuters Brasil, 2021.

WEBB, M. **The Impact of Artificial Intelligence on the Labor Market**. 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3482150. Acesso em: 20 set. 2023.

WINTER, E. **How Many Moves Ahead?** Chess History. 2022. Disponível em: https://www.chesshistory.com/winter/extra/movesahead.html. Acesso em: 20 set. 2023.

ZHONG, H.; XIAO, C.; TU, C.; ZHANG, T.; LIU, Z.; SUNHOW, M. **Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence**, arXiv:2004.12158 [cs]. 18 mai. 2020.