ISSN 2175-6058
DOI: https://doi.org/10.18759/rdgf.v26i2.2235

CONSIDERAÇÕES ÉTICAS NA IMPLEMENTAÇÃO DE SOLUÇÕES BASEADAS EM INTELIGÊNCIA ARTIFICIAL: UMA REVISÃO SISTEMÁTICA DE LITERATURA SOB A PERSPECTIVA JURÍDICA

ETHICAL CONSIDERATIONS IN THE IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE-BASED SOLUTIONS: A SYSTEMATIC REVIEW LITERATURE FROM A LEGAL PERSPECTIVE

Marília Gabriela Silva Lima
Elias Jacob de Menezes Neto

RESUMO

Este artigo realiza uma revisão sistemática da literatura, adotando o método PRISMA, com o propósito de analisar as principais dificuldades na implementação ética de sistemas de inteligência artificial (IA) e identificar os riscos associados ao uso de IA sem parâmetros éticos apropriados. A pesquisa é conduzida em quatro fases: identificação, triagem, elegibilidade e inclusão, consultando artigos nacionais e internacionais nas bases de dados *Scopus* e *Academic Search Premier*, no período de 2017 a novembro de 2022. Dos 163 artigos inicialmente encontrados, 118 foram selecionados para a segunda fase, sendo que 23 foram revisados conforme o escopo do trabalho, resultando na inclusão de 07 artigos nesta revisão. Os resultados evidenciam a urgente necessidade de regulamentação da IA e o estabelecimento de padrões claros para sua implementação, tornando a definição e aplicação dos princípios éticos mais acessíveis e transparentes. Concluímos que sistemas de IA que não seguem parâmetros éticos representam riscos significativos, incluindo a potencial violação de direitos fundamentais e a perpetuação de desigualdades e discriminação. Este estudo destaca a relevância jurídica dessas questões, reforçando a importância de uma abordagem regulatória abrangente para garantir o uso ético e responsável da inteligência artificial.

Palavras-Chave: inteligência artificial; ética; transparência; governança.

ABSTRACT

This article conducts a systematic literature review, employing the PRISMA method, aiming to analyze the main challenges in the ethical implementation of artificial intelligence (AI) systems and identify the risks associated with the use of AI without appropriate ethical parameters. The research is carried out in four phases: identification, screening, eligibility, and inclusion, consulting national and international articles in the Scopus and Academic Search Premier databases from 2017 to November 2022. Out of the initially found 163 articles, 118 were selected for the second phase, with 23 being reviewed according to the scope of the work, resulting in the inclusion of 07 articles in this review. The results highlight the urgent need for AI regulation and the establishment of clear standards for its implementation, making the definition and application of ethical principles more accessible and transparent. We conclude that AI systems not adhering to ethical parameters pose significant risks, including the potential violation of fundamental rights and the perpetuation of inequalities and discrimination. This study emphasizes the legal relevance of these issues, reinforcing the importance of a comprehensive regulatory approach to ensure the ethical and responsible use of artificial intelligence.

Keywords: artificial intelligence; ethics; transparency; governance.

1. INTRODUÇÃO

A inteligência artificial (IA) tem evoluído rapidamente nos últimos anos, com um impacto cada vez maior em diversos campos da sociedade, como direito, medicina, finanças, transporte, entre outros. Para tanto, a IA tem sido amplamente utilizada para solucionar problemas complexos e melhorar a eficiência de processos. Em contrapartida, essa ferramenta tecnológica também é alvo de preocupações por parte dos entusiastas na área, visto que possui o potencial para causar danos sociais e éticos, violando direitos fundamentais dos indivíduos.

Diante dessas preocupações, houve uma crescente necessidade de criar diretrizes éticas para o uso da IA. Ou seja, a implementação da IA deve ser feita observando certos princípios condizentes com uma tecnologia mais responsável. Essas diretrizes são necessárias para garantir que a tecnologia não cause danos à sociedade ou aos indivíduos. Além disso,

as diretrizes éticas também ajudam a garantir que a IA seja desenvolvida e utilizada de forma justa e equitativa.

Portanto, este trabalho trata-se de uma revisão sistemática de literatura seguindo o método de abordagem PRISMA, no qual se busca identificar as principais dificuldades existentes para a implementação de diretrizes éticas nos sistemas de inteligência artificial, bem como quais são os principais riscos de não se utilizar uma IA com esses parâmetros. Além disso, a revisão visa fornecer uma compreensão mais profunda dos desafios e riscos associados à utilização da IA, bem como fornecer informações importantes para futuros trabalhos sobre o assunto.

Assim sendo, este trabalho está dividido em seis seções. A primeira apresenta a introdução do artigo com a contextualização sobre inteligência artificial e aspectos éticos, estabelecendo o campo de abordagem deste estudo. A segunda seção elucida aspectos teóricos sobre inteligência artificial, explicando sobre seu desenvolvimento e possibilidades de atuação, além de esclarecer as diretrizes éticas de inteligência artificial e a necessidade de aplicação. A terceira seção define os procedimentos metodológicos aplicados durante a coleta de dados. Na quarta seção, são apresentadas as fontes de dados utilizadas na pesquisa. Na quinta seção estão elencados os resultados da pesquisa, bem como o compilado das análises. Ao final, a sexta e última seção é destinada às conclusões e considerações finais do estudo científico.

2. CONTEXTO DE DESENVOLVIMENTO DA INTELIGÊNCIA ARTIFICIAL

A tecnologia tem transformado a sociedade de forma profunda. O dinamismo que se verifica entre as ciências, em sentido amplo, tem permitido que soluções inovadoras e versáteis transformem nosso cotidiano. Uma dessas é a inteligência artificial.

Com a produção de dados em padrões nunca antes concebidos, um novo volume processamento de informações pelos computadores ocorreu de forma descomunal. Isso se deu à medida que os dispositivos

se tornaram cada vez mais leves, portáteis e economicamente acessíveis. (Bioni; Luciano, 2019).

Portanto, com a junção desses elementos (maior oferta de dados, maior uso de aplicações possibilitadas pela internet, ampliação da capacidade computacional dos dispositivos e a difusão da tecnologia), foi possibilitada a criação de grandes inovações tecnológicas, feitas a partir de dados e mantendo-se em constante evolução, sendo uma delas a inteligência artificial.

As definições para inteligência artificial (IA) são amplas, de acordo com o contexto de sua atuação. Em uma visão mais abrangente, ela se encontra inserida no ramo da ciência da computação, mas que se ramifica com influência em diversos campos do conhecimento. Segundo Minsky (1988), pode ser definida como a possibilidade de programar computadores para realizar atividades que demandam inteligência quando feitas por humanos. Em outras palavras, a inteligência artificial desenvolve, através de algoritmos programados, sistemas capazes de realizar tarefas que, normalmente, exigiriam o intelecto humano, como aprender, raciocinar, compreender linguagem natural e solucionar problemas.

Uma das técnicas utilizadas pela IA para o cumprimento de sua finalidade denomina-se aprendizado de máquina *(machine learning).* Acerca deste, a Professora Caitlin Mulholland (2019) entende que a técnica conhecida como *machine learning* se configura como qualquer metodologia e conjunto de técnicas que utilizam dados em grande escala para criar conhecimento padrões originais e, com base neles, gerar modelos que são usados para predição a respeito dos dados tratados.

Desta forma, Faceli (2011) esclarece que na aprendizagem de máquina busca-se aprender através de dados, de modo que, a partir destes, a IA possa criar modelos e extrair padrões conforme a necessidade para qual foi criada.

No mais, Barros (2020) destaca que existem diversos tipos de algoritmos de aprendizagem de máquina que estão ganhando maior destaque e maior utilização, principalmente a partir de aplicações complexas com as famosas *Deep Learning.* Entre eles, estão Árvores de Decisão (em inglês*, Decision Tree - DT*), Regressão Logística (em inglês*, Logistic Regression- LR); Random Forest*, *SVM,* Redes Neurais e

outros, mas que para fins cognitivos deste artigo, não cumpre aprofundar acerca de cada um destes, haja vista a suficiência de análise geral sobre a aprendizagem de máquina *(machine learning).*

Voltando os olhos para a necessidade de alimentação dos sistemas através de dados, estes precisam passar por criteriosas análises antes de serem utilizados. Isso porque é possível que os dados venham a refletir padrões preconceituosos e discriminatórios em suas decisões, trazendo consequências negativas à sociedade. Portanto, é importante garantir que os dados da IA sejam representativos e não estejam influenciados por preconceitos ou julgamentos errôneos.

Conforme exposto anteriormente, para que se obtenha um resultado favorável com a utilização de modelos criados pela inteligência artificial, é necessário garantir a qualidade dos dados utilizados pelos algoritmos de aprendizagem. Portanto, é fundamental ressaltar que é responsabilidade humana "alimentar" a máquina, para que a partir deles, ela possa propor soluções e apresentar resultados. (Faceli, 2011).

Além disso, dependendo da qualidade, quantidade e diversidade dos dados coletados pela máquina, os resultados obtidos através de sua utilização podem infringir direitos fundamentais dos seres humanos. É possível identificar inúmeros casos já ocorridos, em que a máquina reproduziu ações e erros grotescos da sociedade.

Um desses exemplos foi ocasionado por um dos gigantes da tecnologia: A empresa Google. Um usuário realizou o download de algumas fotos tiradas junto com sua amiga, e lamentavelmente, o algoritmo utilizado pelo sistema alocou a foto da usuária em um álbum nomeado de "gorilas". Isto ocorreu pelo fato de as usuárias serem negras, e o algoritmo repetiu um comportamento racista inerente à sociedade, que seria o preconceito, fazendo com que esta prática fosse repassada para o sistema de inteligência artificial. (Harada, 2015).

Na sua manifestação, o Google desculpou-se e declarou que estaria trabalhando para que essas práticas não se repetissem. Cumpre salientar que neste ato foi identificada claramente a ofensa aos direitos fundamentais das usuárias, como por exemplo, a honra e a imagem, estando estes previstos em nossa constituição federal.

Diante da existência de situações deste tipo, a comunidade acadêmica e científica passou a focar em formas de inibir ou controlar esse tipo de acontecimento. Algumas das principais formas de mudanças consistem em coletar fontes diversas e representativas para garantir que o modelo de IA seja alimentado por uma ampla variedade de perspectivas e informações, ou ainda, realizar análises de risco dos dados utilizados e implementar medidas de transparência, buscando garantir que as decisões baseadas em IA sejam claras e compreensíveis para todos. (Turing, 2019).

Conforme será demonstrado no próximo tópico, esses esforços para a redução de problemas causados pelo uso de IA encontram-se fundamentados em diretrizes éticas. Diante do aumento substancial do uso da inteligência artificial no mundo contemporâneo, em diferentes áreas, há a possibilidade de causar impacto sociológico em uma parcela considerável de indivíduos.

2.1. DIRETRIZES ÉTICAS DE INTELIGÊNCIA ARTIFICIAL E SUA NECESSIDADE DE APLICAÇÃO

Um dos primeiros autores a discutir diretrizes ou leis para robôs e sistemas de inteligência foi Isaac Asimov. Em seu conto intitulado de "Runaround" de 1942 (posteriormente publicado no livro "I Robot"), ele criou as conhecidas Três Leis da Robótica, que inspiraram muitos pesquisadores da área e foram fundamentais para as diretrizes éticas projetadas posteriormente para a IA. São elas: Um robô não pode ferir um ser humano ou, por omissão, permitir que um ser humano venha a ferir; Um robô deve obedecer às ordens dadas por seres humanos, exceto quando tais ordens entrarem em conflito com a Primeira Lei; e por fim, um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a Primeira ou Segunda Lei. (Asimov, 2004).

Após décadas de pesquisa no campo da IA, dilemas surgidos no mundo da ficção científica avançaram para o mundo real. Essas interações do humano com as aplicações de IA foram recebidas algumas vezes de forma positiva, mas, às vezes, também incitando preocupações e medos

entre a população humana, incluindo especialistas em IA. (Johnson; Verdicchio, 2017).

Portanto, com a inteligência artificial adentrando em diversas esferas da vida íntima do ser humano, isso indica que problemas éticos também começaram a existir, conforme já ressaltado neste estudo. Para tanto, alguns destes dilemas já foram objetos de análise, por exemplo, quando um carro autônomo se encontra em posição de decisão e precisar escolher entre salvar a vida de três pessoas fora do carro em vez da vida de um motorista. (Bonnefon, Shariff; Rahwan, 2016)

Destarte, fica o questionamento acerca da situação: Existem conjuntos de valores predefinidos ou com base em que o carro deve tomar a decisão?

Conforme elucidado em tópico anterior, as aplicações de IA são alimentadas por meio de dados, podendo esses dados conter vieses em sua composição, haja vista que provêm de informações produzidas por seres humanos. Pois bem, a partir disso, quando o carro precisa tomar a decisão sozinho, também pode haver um viés existente no algoritmo do carro, acionado por um fator humano. (Wang; Siau, 2018).

Portanto, são inúmeros os casos que podemos perceber ser primordial a inclusão de diretrizes éticas na aplicação da IA, utilizando conceitos éticos desde a sua concepção e adentrando em todas as cadeias de produção da inteligência artificial. Algumas dessas diretrizes já foram abordadas por diferentes entidades públicas ou privadas, tais como: OECD (Organização para a Cooperação e Desenvolvimento Econômico), TURING INSTITUTE (Alan Turing Institute é o instituto nacional de ciência de dados e inteligência artificial do Reino Unido), e a OAI- UK. No âmbito brasileiro, o Conselho Nacional de Justiça e o Tribunal de Contas da União são alguns dos órgãos que estão angariando esforços no campo da IA e ética.

Como resultado de pesquisas já realizadas nessa área, os estudiosos elencaram várias diretrizes éticas que deveriam ser seguidas no momento da implementação da inteligência artificial. Para este estudo, resolvemos focar nas de maior importância para a análise realizada bibliograficamente, sendo elas: Transparência, responsabilidade, e não discriminação.

A transparência é uma das diretrizes éticas mais importantes na implementação da inteligência artificial. Ela diz respeito à capacidade das

pessoas compreenderem como a IA está tomando decisões e quais são as informações e fatores para que a IA ofereça determinado resultado. Além disso, a transparência é importante porque permite que as pessoas adquiram uma maior confiança na IA e nos seus resultados, entendendo também o impacto que ela causa. Sobre este ponto, vale salientar que, através da transparência é possível um auxílio a mais para identificar e corrigir erros e problemas éticos que possam ocorrer, garantindo que a IA seja usada de forma responsável e ética. (Rauber; Trasarti; Giannotti, 2019).

De acordo com Leslie (2019), a responsabilidade, chamada de *accountability* em inglês, engloba principalmente a possibilidade prestação de contas, na qual deve ser exigido aos criadores e a todos aqueles que se dispõem a fazer uso de sistemas inteligentes. Portanto, é necessário que exista a possibilidade de se auditar o sistema de inteligência artificial, trazendo à tona a possibilidade de obter explicações e justificativas sobre os fatores determinantes para o processo decisório do algoritmo, assim como das etapas do seu desenvolvimento.

Com relação ao princípio da não discriminação, este estabelece que a IA não deve ser utilizada de forma discriminatória, ou seja, deve evitar ou corrigir o preconceito ou a discriminação em relação a raça, gênero, orientação sexual, identidade de gênero, idade, capacidade ou qualquer outra característica protegida por lei. É uma importante diretriz a ser analisada, visto que as consequências de decisões discriminatórias perpassam pelas esferas mais intrínsecas do ser humano, ressaltando gatilhos indesejados. (Mehabi, 2021).

Diante do exposto, partimos do princípio de que a aplicação de diretrizes éticas na implementação de uma inteligência artificial é fundamental para a garantia e resguardo de diversos direitos fundamentais dos indivíduos. Além disso, neste estudo, buscamos entender as principais dificuldades que os acadêmicos relatam para a produção de uma IA ética e responsável, bem como quais seriam os principais riscos da utilização desses sistemas sem seguir as diretrizes éticas. Na próxima seção iremos apresentar o método escolhido para o referido artigo, as questões de pesquisa que irão nortear o trabalho e as estratégias de busca nas bases bibliográficas escolhidas.

3. PROCESSOS METODOLÓGICOS

Para a realização desta revisão, a pesquisa bibliográfica partiu dos seguintes questionamentos: "Quais são as principais dificuldades para utilização de IA que sigam balizas éticas?" e "Quais os riscos de se utilizar uma IA que não segue diretrizes éticas?".

As revisões sistemáticas baseiam-se em perguntas claras, utilizando métodos sistematizados e explícitos com objetivo de identificar, selecionar e avaliar criticamente pesquisas relevantes. Nesse sentido, optou-se pela utilização da recomendação PRISMA, com as devidas adaptações para o estudo em questão, visando auxiliar autores a melhorarem a qualidade de suas revisões sistemáticas e metanálises.

Como os estudos avaliados apresentaram características diversas, incluindo campo de atuação, objetivos e procedimentos metodológicos heterogêneos, não foi possível a sua metanálise. Porém, após análise dos dados, o levantamento possibilitou o estabelecimento de considerações acerca das principais dificuldades para implementação de soluções de IA éticas e dos principais riscos de se utilizar uma IA que não segue diretrizes éticas.

4. FONTES DE DADOS

Foi realizada busca nas bases de dados *Scopus e Academic Search Premier*, publicados no período de 2017 a 2022[1], conforme ilustrado através da figura 02, indexado no corpo deste trabalho. Para a pesquisa foram utilizados descritores — palavras-chave para recuperação dos assuntos na literatura. Os cruzamentos desses descritores foram realizados nos idiomas inglês e português, no qual formaram a seguinte string de busca: *("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Inteligência artificial") AND ("Judicial Power" OR "Legal" OR "Judge" OR "Justice" OR "judiciary" OR "government" OR "Judiciário") AND ("Moral" OR "Rules" OR "Regulations OR ethics" OR "ethical" OR "Ética") AND ("Human rights" OR "individual rights" OF "fundamental rights" OR "lesão aos direitos fundamentais").*

Para a composição desta Revisão Sistemática de Literatura - RSL, estabelecemos alguns critérios que foram divididos em duas categorias: CI) Critérios de Inclusão e CE) Critérios de Exclusão.

Quadro 01- Critérios de inclusão e exclusão

Critérios de inclusão (CI)	Critérios de exclusão (CE)
CI 1- Artigos em Língua portuguesa, língua inglesa e língua espanhola.	CE 1- Artigos não apresentam algum método, técnica, revisão de literatura, modelo ou ferramenta para utilização de IA ética e responsável, e/ou que demonstrem a ameaça de uma IA aos direitos fundamentais implementada sem requisitos éticos.
CI 2-Artigos completos que demonstrem dificuldades e necessidades para implementação da IA responsável; e artigos que demonstrem a ameaça de uma IA aos direitos fundamentais implementada sem requisitos éticos.	CE 2- Artigos que não se relacionam com as questões de pesquisa.
CI 3- Artigos completos com resultados consistentes com os objetivos de pesquisa.	CE 3- Artigos em idiomas diferentes de português, inglês e espanhol.
CI 4- Trabalhos publicados no período de 2017 a 2022 (últimos 5 anos).	CE 4- Artigos duplicados
	C5- Artigos de revisão de literatura

Fonte: Elaborado pelos autores (2022).

Os artigos foram selecionados a partir da utilização dos descritores da *string* de busca definida e a identificação foi realizada em três etapas, a saber: Etapa 1: Pesquisa dos artigos nas bases de dados, com a seleção dos artigos através do título e exclusão de estudos duplicados; Etapa 2: leitura dos resumos e conclusões dos estudos selecionados na etapa 1 e exclusão daqueles que não se adequaram aos critérios de inclusão; Etapa

3: leitura na íntegra de todos os estudos restantes das etapas anteriores e seleção dos que se enquadraram nos critérios de inclusão, por meio de protocolo criado para esse fim.

Inicialmente na primeira fase, foram obtidos 163 artigos oriundos das bases de dados pesquisadas *(Scopus e Academic Search Premier*). Destes, 86 vieram da *Scopus* e 77 vieram da *Academic Search Premier.* Ressalta-se que na base de dados *Academic*, dos 77 artigos selecionados 27 encontravam-se duplicados na própria base, sendo, portanto, excluídos pela própria plataforma, restando um total de 136 artigos analisados. Destes 136, após unificar os artigos encontrados na *Scopus* e na *Academic*, ainda identificamos mais 35 artigos duplicados. Após o saneamento da duplicação, restou um total de 45 artigos duplicados e removidos, e 118 artigos escolhidos para a segunda fase.

Sendo assim, na segunda fase, após a análise dos títulos dos artigos, foram selecionados para a terceira fase 12 artigos oriundos da *Scopus*, e 11 artigos oriundos da *Academic Search Premier.* Para a terceira fase, após a análise dos resumos e conclusões, restaram para a leitura do texto completo 6 artigos vindos da *Academic* e mais 4 artigos vindos da *Scopus*. Após a leitura do texto completo e conforme os critérios de elegibilidade, foram selecionados 7 artigos para esta revisão. Para tanto, criamos a figura 01 como forma de ilustrar os dados aqui apresentados.

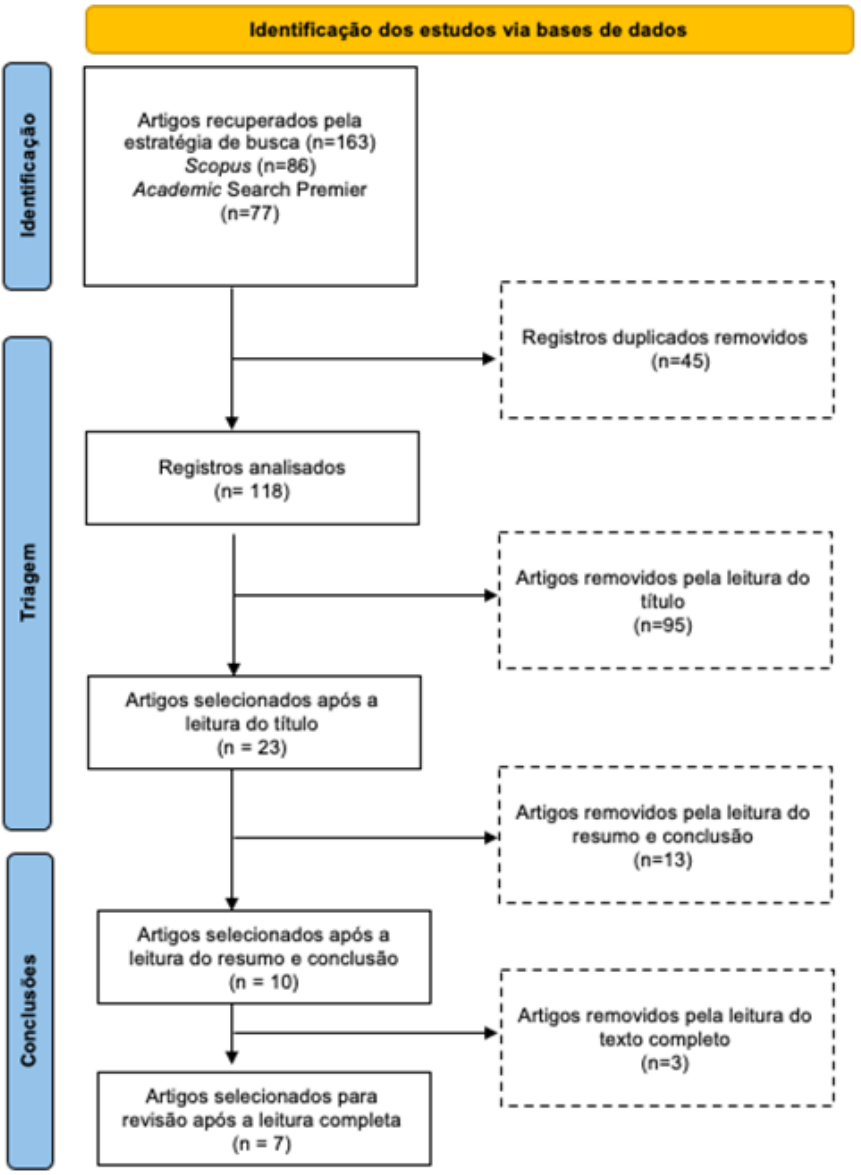

Figura 01 - Fluxograma do número de artigos encontrados e selecionados após aplicação dos critérios de inclusão e exclusão.

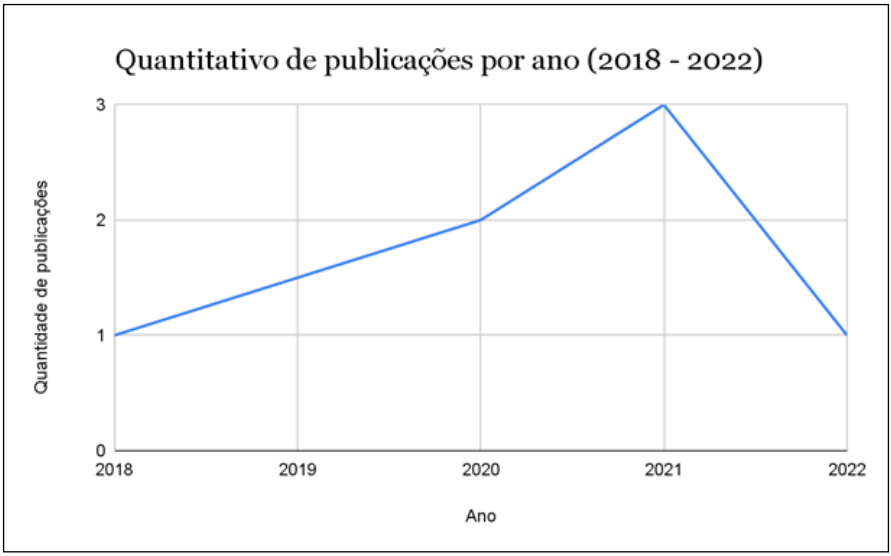

Figura 02 - Quantitativo de publicações por ano. Para o ano de 2022, foram considerados os artigos publicados até o dia 01 de novembro.

5. RESULTADOS E DISCUSSÕES

No quadro 02, tem-se a listagem dos sete trabalhos que foram mantidos para o mapeamento sistemático. Sendo assim, organizamos a tabela através dos identificadores [T01 a T07], sinalizando o título do trabalho, autores, ano de publicação e o tipo.

Quadro 02. Lista dos trabalhos selecionados:

ID	Título	Autor	Ano	Tipo
T01	AI Challenges and the Inadequacy of Human Rights Protections.	LIU, Hin-Yan.	2021	Artigo
T02	Artificial intelligence and human rights: Between law and ethics	SARTOR, Giovanni.	2020	Artigo

T03	Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines	FUKUDAPARR, Sakiko; GIBBONS, Elizabeth.	2021	Artigo
T04	From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence.	LIU, Hin-Yan; ZAWIESKA, Karolina.	2020	Artigo
T05	Governing artificial intelligence: ethical, legal and technical opportunities and challenges.	CATH, Corinne.	2018	Artigo
T06	International Human Rights, Artificial Intelligence, and the Challenge for the Pondering State: Time to Regulate?	BELLO Y VILLARINO, José-Miguel; VIJEYARASA, Ramona.	2022	Artigo
T07	Legal regulation of the use of artificial intelligence: problems and development prospects	Yara O., Brazheyev A., Golovko L., Bashkatova V.	2021	Artigo

Fonte: Elaborado pelos autores (2022).

Dentre os autores citados, não há nenhum brasileiro na lista e uma forte concentração de publicações sobre o tema nos Estados Unidos e Europa. A falta de publicações nacionais dificulta a comparação de cenários, já que as situações políticas e socioeconômicas são diferentes das regiões citadas e afetam diretamente o desenvolvimento de tecnologias de inteligência artificial, e consequentemente de normativas a respeito do tema.

Na sequência, iremos analisar as ideias centrais discutidas por cada trabalho selecionado, possibilitando uma visão panorâmica das produções selecionadas para este estudo.

Liu, Hin-Yan (2021) realizou um estudo com o objetivo de demonstrar alguns dos desafios existentes nas aplicações de inteligência artificial (IA) para a proteção e gozo dos direitos humanos. Afirma que, ao seu ver, os principais problemas relacionados à inteligência artificial ainda são desconhecidos, e que poderão gerar resultados inesperados ou, ainda, ser a fonte de uma sensação persistente de que existem inadequações e deficiências no regime de direitos humanos existente. Afirma ainda que, para enfrentar os desafios impostos pela IA, são necessárias pesquisas ativas, sistemáticas e contínuas dos potenciais problemas, que precisam ser conduzidas tanto para tentar atender ao problema em si, quanto para tentar identificá-lo precocemente.

Sartor, Giovanni (2020), examina a maneira como a IA é abordada por regras, princípios e argumentos éticos e legais. Considera até que ponto as demandas do direito e da ética podem se sobrepor e examina como elas podem ser coordenadas. No particular, argumenta que os direitos humanos/fundamentais e os valores sociais são centrais tanto para a ética quanto para o direito. Embora possam ser enquadrados de diferentes maneiras, eles podem fornecer uma referência normativa útil para vincular ética e direito ao abordar as questões normativas que surgem em conexão com a IA.

Fukuda-Parr e Sakiko; Gibbons, Elizabeth (2021), analisaram, em seu estudo, 15 diretrizes pré-selecionadas que seriam mais fortes em direitos humanos e saúde global, tendo por base princípios de direitos humanos (igualdade, participação e responsabilidade), bem como o direito à privacidade. Afirmam que algumas das diretrizes éticas atualmente utilizadas possuem dificuldades para lidar com desigualdades e discriminação. Argumentaram, ainda, a necessidade de criar um conjunto de "normas de fato" assim como reinterpretar a noção de direitos humanos de modo a abarcar critérios éticos no campo da inteligência artificial.

Liu, Hin-Yan e Zawieska, Karolina (2020), apresentaram em seu estudo a necessidade de impulsionar o uso da inteligência artificial de modo responsável, propondo um conjunto complementar de direitos

humanos, direcionado especificamente contra os danos decorrentes das tecnologias robóticas e de inteligência artificial (IA).

Cath, Corinne (2018), apresenta em seu estudo análises sobre questões éticas, desafios jurídico-regulatórios e técnicos colocados pelo desenvolvimento de regimes de governança para sistemas de IA. Ela também fornece uma visão geral dos desenvolvimentos recentes na governança da IA, com discussões acerca da regulamentação da IA, estruturas éticas e abordagens técnicas, além de fornecer algumas sugestões concretas para aprofundar o debate sobre a governança da IA.

Bello Y Villarino, José-Miguel e Vijeyarasa, Ramona (2022), analisaram os riscos e vantagens da regulamentação precoce da inteligência artificial (IA) do ponto de vista dos direitos humanos internacionais. Ao explorar argumentos de documentos acadêmicos e políticos de várias jurisdições sobre possíveis abordagens para regulamentar a IA, os autores identificam uma tendência atual entre os estados de esperar em vez de regular proativamente.

Por fim, Yara O., Brazheyev A., e Golovko L., Bashkatova V (2021), identificaram perspectivas para o uso da IA no campo jurídico, no qual elucidaram a necessidade de regulamentação legal do uso de IA, junto com um código de ética para inteligência artificial que evite sua má aplicação e minimize possíveis consequências prejudiciais.

Nesse contexto, a seção a seguir apresenta os resultados qualitativos da análise dos 07 artigos finais selecionados, considerando-se as duas questões de pesquisa estabelecidas previamente e que tinham como objetivo: O estabelecimento de considerações acerca das principais dificuldades para implementação de soluções de IA éticas e os principais riscos de se utilizar uma IA que não segue diretrizes éticas.

5.2. ANÁLISE QUALITATIVA DOS ESTUDOS

QP1- Quais são as principais dificuldades para utilização de IA que sigam balizas éticas?

O estudo [T01] demonstra que um dos principais desafios da utilização de IA com balizas éticas envolve o conceito de direitos humanos. Para

ele, não existe um rol previamente definido desses direitos, que devem ser compreendidos como uma construção do ordenamento jurídico, não como algo dado. Assim, a dificuldade do sistema jurídico em reconhecer as violações de Direitos Humanos potencialmente causadas por IA pode levar a uma situação de vazio normativo, sem proteção do ordenamento jurídico.

Portanto, se o regime de direitos humanos não consegue identificar e reconhecer os desafios colocados pela IA, os danos e os sofrimentos decorrentes do seu uso serão condicionados a este entendimento. O texto afirma que a própria interação dos seres humanos com a inteligência artificial é capaz de influenciar o processo de tomada de decisões humanas. Com isso, gera-se uma modificação na valoração de direitos humanos de uma maneira que não é facilmente perceptível. Portanto, para ele, outro desafio seria a necessidade de identificação precoce dessas violações causada pela inteligência artificial, antes mesmo de um reconhecimento por meio dos órgãos ou instituições que zelam pelos direitos humanos.

Por fim, o texto conclui que, para enfrentar os desafios impostos pela IA, são necessárias pesquisas ativas, sistemáticas e contínuas dos potenciais problemas advindos da ausência de balizas éticas em seu uso. Isso seria válido tanto para tentar atender ao problema de informação quanto para identificar problemas antes que eles se transformem em dificuldades maiores.

Nesse ponto, vale salientar a divergência parcial do exposto no trabalho [T01] com o trabalho [T03]. Os autores da pesquisa [T03] afirmam haver uma diferenciação entre as abordagens éticas e abordagem de direitos humanos, a depender da maneira pela qual a responsabilidade e as assimetrias de poder são abordadas. Alegam, ainda, que os direitos humanos são estruturados não apenas como um conjunto de valores, mas, também, para funcionar como um controle sobre o poder arbitrário.

Retornando a análise para a sequência disposta no quadro 01, o trabalho [T02] demonstra que uma das principais dificuldades para implementação de uma IA ética consiste na necessidade de equilibrar direitos e deveres relativos à implantação de aplicações de IA, que podem trazer consequências tanto no campo da ética quanto na lei. Afirma que pode existir uma tensão entre a implantação da IA e os direitos individuais (e valores sociais), como, por exemplo, o aumento da eficiência econômica

impulsionados pela IA e os impactos negativos na autonomia e igualdade de trabalhadores homens e mulheres.

Os pesquisadores da [T03] apontaram que a própria implementação de soluções de IA éticas são os desafios mais críticos. Alegam que ocorre uma divergência significativa em como os princípios éticos são interpretados, por que eles são considerados importantes, e como eles devem ser implementados. Acrescentam, ainda, que existe uma necessidade latente de desenvolvimento de fortes mecanismos regulatórios.

Na pesquisa demonstra-se a importância de um maior esclarecimento acerca dos princípios éticos de IA e sugerem que sejam realizadas definições universais sobre o tema em questão. Isso para que sejam aplicadas a todos aqueles que trabalhem com implementação de inteligência artificial, construindo uma espécie de documento que norteia de forma global os sistemas de IA.

O estudo [T05] demonstra que a ausência de governança sobre a IA é um dos principais desafios para a sua utilização de forma ética. Fica claro na pesquisa que os autores demonstram insatisfação com o estado de governança de IA à época que foi escrito o trabalho. Argumentam, ainda, que a ausência de regulamentação abre margem para que as grandes empresas de tecnologia que tratam dados pessoais utilizem-nos de forma indiscriminada. Ressaltam que é necessária uma vigilância permanente para evitar uma "regulamentação americanizada", já que é justamente naquele país que se encontram as gigantes da tecnologia de inteligência artificial, o que amplia sua capacidade de direcionar o debate público e regulatório sobre o tema.

A pesquisa [T06] não se atém a apenas um desafio de implementação de IA ética, e sim, direciona seu olhar em algumas outras vertentes. De início os autores fazem apontamentos sobre os desafios de utilização de IA, sendo possível perceber que a [T06] converge em partes com o que foi exposto na [T02]. Segundo os autores da [T06], é necessário não apenas equilibrar custos e benefícios para os seres humanos do uso de IA, mas expandir as proteções de direitos para sistemas inteligentes. O debate, nesse caso, é principalmente ético e vinculado à inteligência e a autonomia crescentes dos sistemas de IA. Para [T06], o problema é que o regime de direitos humanos deve ser estendido e aplicado aos próprios

sistemas quando eles atingem um nível de autonomia que envolve uma espécie de status moral.

Em segundo ponto, os autores da [T06] concordam com o abordado pelas pesquisas [T03] e [T05], especialmente sob a ótica de que a ausência de regulamentação da inteligência artificial é uma das principais dificuldades para que se possa utilizar IA calcada em princípios éticos. Eles justificam que a necessidade de lidar com esses riscos requer urgentemente a intervenção do Estado. Além disso, os autores alegam que a ausência de atividade regulatória é uma falha grave do Estado, visto que a omissão na definição de normas juridicamente vinculantes para proteger os direitos humanos em face dos sistemas de IA – é, por si só, uma violação das normas do direito internacional de direitos humanos.

Nesse sentido, afirmam que é indispensável uma legislação capaz de ir além de uma mera implementação de políticas, elevando-se ao nível de uma obrigação do Estado sob o regime de sistema internacional de direitos humanos.

QP2- Quais os riscos de se utilizar uma IA que não segue diretrizes éticas?

Sobre o segundo questionamento, a pesquisa [T07] abordou o tema de forma contundente. Os autores citam alguns casos de utilização da IA no sistema judiciário, como, por exemplo, o software "Cassandra" com elementos de IA que permite analisar a possibilidade de reincidência em matéria penal.

Para os autores, é necessário o desenvolvimento de um Código de Ética para inteligência artificial e legislação que evite sua má aplicação e minimize possíveis consequências prejudiciais, como a discriminação de grupos vulneráveis da sociedade, como pessoas com deficiências, imigrantes, mulheres e minorias étnicas e raciais.

A pesquisa [T02] afirma que a ausência de balizas éticas relacionadas à IA pode trazer riscos em diversos campos de atuação, como na esfera da responsabilidade civil, seguros, proteção de dados, segurança, contratos e crimes, trazendo alguns exemplos no corpo da pesquisa para demonstrar suas afirmações.

Os autores ainda afirmam que, para implementar e alcançar uma IA confiável, alguns requisitos precisam ser atendidos com base nos seguintes princípios: Agência e supervisão humana, incluindo direitos fundamentais; Robustez e segurança técnica, incluindo resiliência a ataques e segurança, planos de recuo e segurança geral, precisão, confiabilidade e reprodutibilidade; Privacidade e governança de dados, incluindo respeito pela privacidade, qualidade e integridade dos dados e acesso aos dados; Transparência, incluindo rastreabilidade, explicabilidade e comunicação; Diversidade, não discriminação e justiça, incluindo evitar preconceitos injustos, acessibilidade e design universal, e participação das partes interessadas; Bem-estar social e ambiental, incluindo sustentabilidade e respeito pelo ambiente, impacto social, sociedade e democracia; Responsabilidade, incluindo auditabilidade, minimização e relato de impactos negativos.

Em consonância com o exposto na [T07] os autores da [T03] alegaram a necessidade de análise dos conjuntos de dados utilizados no treinamento dos modelos de IA, visto que estes podem refletir os preconceitos existentes na sociedade, que discriminam grupos específicos. Portanto, a ausência de diretrizes éticas para implementação de IA, torna o indivíduo suscetível a situações de discriminação. Os autores afirmam, ainda, que os danos da discriminação devem ser identificados desde o início do sistema, até completar todo o seu ciclo de atuação, e alegam que os Estados devem documentar as ações tomadas para mitigar o tratamento desigual entre pessoas.

A pesquisa [T04] aborda esse problema ao discorrer sobre a necessidade de responsabilização das decisões tomadas pela IA. Os autores afirmam ser necessário as diretrizes éticas para que ocorra uma devida regulamentação da responsabilização, e, assim, poder punir e apurar eventuais problemas que venham a ser ocasionados pelo uso de IA. Afirmam, ainda, que a responsabilidade no uso de IA é necessária para estabelecer as esferas de obrigação para os envolvidos no projeto e desenvolvimento da robótica e fixar a responsabilidade pelo uso dessas tecnologias.

A partir da análise dos artigos finais selecionados, é evidente que ainda há muito a ser estudado e compreendido sobre as diretrizes éticas da

IA e sua implementação em sistemas práticos, bem como as dificuldades de sua utilização. As questões éticas e sociais relacionadas à IA são complexas e multifacetadas, no qual requerem uma abordagem interdisciplinar para serem abordadas adequadamente.

Além disso, a falta de uma regulamentação clara e de consenso sobre os princípios éticos da IA representa um desafio para sua implementação eficaz. É crucial continuar a explorar as dificuldades e desafios para a implementação de sistemas de IA éticos e responsáveis, bem como os riscos de se utilizar sistemas de IA que não seguem princípios éticos, a fim de garantir que a tecnologia seja utilizada de maneira justa e responsável.

6. CONSIDERAÇÕES FINAIS

A revisão sistemática de literatura realizada proporcionou uma análise aprofundada sobre as principais dificuldades para a utilização e implementação ética da inteligência artificial (IA) e os riscos associados ao uso de IA que não segue diretrizes éticas. Os estudos analisados forneceram insights valiosos, revelando uma série de desafios e preocupações que permeiam esse campo em constante evolução.

Dentre os resultados obtidos, pode-se destacar a complexidade associada à integração de balizas éticas na IA, especificamente no contexto dos direitos fundamentais. A inexistência de um conjunto pré-definido de direitos para orientar a IA, conforme argumentado no [T01], sugere uma necessidade crítica de clarificação e reconhecimento desses direitos pelo sistema jurídico. A dificuldade do sistema jurídico em identificar e reconhecer violações de direitos humanos causadas pela IA cria um potencial vazio normativo, expondo os usuários a danos e sofrimentos sem proteção legal. A interação entre seres humanos e IA pode influenciar sutilmente a tomada de decisões humanas, apresentando um desafio adicional na valoração dos direitos fundamentais dos indivíduos.

Portanto, enfrentar os desafios da IA requer pesquisas ativas e sistemáticas para compreender e mitigar os problemas decorrentes da falta de balizas éticas. A identificação precoce de violações éticas

pela IA é destacada nos estudos como uma prioridade para evitar consequências prejudiciais.

Os estudos ainda demonstraram os riscos concretos da utilização de IA que não segue balizas éticas, citando exemplos no sistema judiciário, como o software "Cassandra". Essa IA pode resultar em discriminação contra grupos vulneráveis da sociedade, exigindo a necessidade de um Código de Ética para mitigar tais consequências. Além disso, destacou-se nos artigos riscos em campos diversos, como responsabilidade civil, seguros, proteção de dados, segurança, contratos e crimes, enfatizando a necessidade de balizas éticas em cada um desses domínios.

Desse modo, a revisão sistemática demonstrou a importância de diretrizes éticas para a responsabilização eficaz das decisões tomadas pela IA, sublinhando a necessidade de regulamentação para punir e apurar problemas decorrentes do uso inadequado de IA. Isso porque a análise geral dos estudos revela a complexidade das questões éticas e sociais relacionadas à IA. A falta de uma regulamentação clara e consensual sobre os princípios éticos da IA representa um desafio para sua implementação eficaz. Portanto, é imperativo continuar explorando as dificuldades e desafios na implementação de sistemas de IA éticos e responsáveis, garantindo que a tecnologia seja utilizada de maneira justa e responsável.

NOTA

[1] Para o ano de 2022, foram considerados os artigos publicados até o dia 01 de novembro.

REFERÊNCIAS

ASIMOV, Isaac. I, Robot. **The Robot Series**. 2004.

BARROS, Thiago Medeiros. **Um processo orientado a dados para geração de modelo de predição de evasão escolar**. 2020. 116f. Tese (Doutorado em Engenharia Elétrica e de Computação) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2020. Disponível em: https://repositorio.ufrn.br/handle/123456789/31933. Acesso em: 1º fev. 2023.

BELLO Y VILLARINO, José-Miguel; VIJEYARASA, Ramona. International Human Rights, Artificial Intelligence, and the Challenge for the Pondering State: Time to Regulate?. **Nordic Journal of Human Rights**, p. 1-22, 2022. Disponível em: https://www.tandfonline.com/doi/full/10.1080/18918131.2022.2069919. Acesso em: 04 nov. 2022.

BIONI, Bruno Ricardo; LUCIANO, Maria. O Princípio da Precaução na Regulação de Inteligência Artificial: seriam as leis de proteção de dados o seu portal de entrada. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coor.). **Inteligência artificial e direito**: ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. Disponível em: https://brunobioni.com.br/wp-content/uploads/2019/09/Bioni-Luciano_O-PRINCI%CC%81PIO-DA-PRECAUC%CC%A7A%CC%83O-PARA-REGULAC%CC%A7A%CC%83O-DE-INTELIGE%CC%82NCIA-ARTIFICIAL-1.pdf. Acesso em: 24 nov. 2022.

BONNEFON, Jean-François; SHARIFF, Azim; RAHWAN, Iyad. The social dilemma of autonomous vehicles. **Science**, v. 352, n. 6293, p. 1573-1576, 2016. Disponível em: https://www.science.org/doi/abs/10.1126/science.aaf2654. Acesso em: 4 fev. 2023.

CATH, Corinne. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 376, n. 2133, p. 20180080, 2018. Disponível em: https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0080. Acesso em: 05 dez. 2022.

FACELI, K. et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. São Paulo: Editora LTC, 2011.

FUKUDA-PARR, Sakiko; GIBBONS, Elizabeth. Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines. **Global Policy**, v.

12, p. 32-44, 2021. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1111/1758-5899.12965. Acesso em: 8 nov. 2022.

JOHNSON, Deborah G.; VERDICCHIO, Mario. AI anxiety. **Journal of the Association for Information Science and Technology**, v. 68, n. 9, p. 2267-2270, 2017. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23867. Acesso em: 1º fev. 2023.

LESLIE, David. **Understanding artificial intelligence ethics and safety**, 2019. Disponível em: https://arxiv.org/pdf/1906.05684.pdf. Acesso em: 2 fev. 2023.

LIU, Hin-Yan. AI Challenges and the Inadequacy of Human Rights Protections. **Criminal Justice Ethics**, v. 40, n. 1, p. 2-22, 2021. Disponível em: https://www.tandfonline.com/doi/abs/10.1080/0731129X.2021.1903709. Acesso em: 1º nov. 2022.

LIU, Hin-Yan; ZAWIESKA, Karolina. From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence. **Ethics and Information Technology**, v. 22, n. 4, p. 321-333, 2020. Disponível em: https://link.springer.com/article/10.1007/s10676-017-9443-3. Acesso em: 7 nov. 2022.

MEHRABI, Ninareh et al. A survey on bias and fairness in machine learning. **ACM Computing Surveys (CSUR)**, v. 54, n. 6, p. 1-35, 2021. Disponível em: https://dl.acm.org/doi/abs/10.1145/3457607. Acesso em: 6 fev de 2023.

MINSKY, Marvin. **Society of mind**. Simon and Schuster, 1988.

MULHOLLAND, Caitlin. Responsabilidade civil e processos decisórios autônomos em sistemas de Inteligência Artificial (IA): autonomia, imputabilidade e responsabilidade. **Inteligência artificial e direito: ética, regulação e responsabilidade. Editora Revista dos Tribunais**, 2019.

RAUBER, Andreas; TRASARTI, Roberto; GIANNOTTI, Fosca. Transparency in algorithmic decision making. **ERCIM News**, v. 116, p. 10-11, 2019.Disponível em: https://ercim-news.ercim.eu/images/stories/EN116/EN116-web.pdf#page=10. Acesso em: 22 jan. 2023.

SARTOR, Giovanni. Artificial intelligence and human rights: Between law and ethics. **Maastricht Journal of European and Comparative Law**, v. 27,

n. 6, p. 705-719, 2020. Disponível em: https://journals.sagepub.com/doi/abs/10.1177/1023263X20981566. Acesso em: 04 nov. 2022.

TURING, Allan M. **Computing Machinery and Intelligence**. Mind, Oxford, v. 59, n.236, p.443-460, out. 1950. Disponível em: http://www.jstor.org/stable/2251299. Acesso em: 14 jan. 2023.

WANG, Weiyu; SIAU, Keng. Ethical and Moral Issues with AI - A Case Study on Healthcare Robots. Missouri: **Emergent Research Forum (ERF)**, 2018. Disponível em: https://www.researchgate.net/profile/Keng-Siau-2/publication/325934375_Ethical_and_Moral_Issues_with_AI/links/5b97316d92851c78c418f7e4/Ethical-and-Moral-Issues-with-AI.pdf. Acesso em: 2 fev. 2023.

YARA, Olena et al. Legal regulation of the use of artificial intelligence: Problems and development prospects. **European Journal of Sustainable Development**, v. 10, n. 1, p. 281-281, 2021. Disponível em: http://www.ecsdev.org/ojs/index.php/ejsd/article/view/1170. Acesso em: 4 nov. 2022.

Recebido em:10-2-2023
Aprovado em:29-12-2025

Marília Gabriela Silva Lima

Mestranda em Direito pela Universidade Federal do Rio Grande do Norte. Pós graduada em Propriedade intelectual pela Universidade Norte do Paraná - UNOPAR. Pós graduada em Direito Digital e Proteção de Dados pela Escola Brasileira de Direito - EBRADI. Advogada OAB/RN. Membro do Grupo de Pesquisa em Inovação, Direito e Novas Tecnologias do Ministério Público do Paraná. E-mail: mariiliagsilvalima@hotmail.com - Lattes: http://lattes.cnpq.br/1063447019412961

Elias Jacob de Menezes Neto

Doutor em Direito. Professor da área de Aprendizado de Máquina do Instituto Metrópole Digital da Universidade Federal do Rio Grande do Norte. Bolsista de Produtividade em Desenvolvimento Tecnológico e Extensão Inovadora do CNPq. E-mail: elias.jacob@ufrn.br - Lattes: http://lattes.cnpq.br/9152955193794784

Universidade Federal do Rio Grande do Norte

Avenida Senador Salgado Filho, s/n, Lagoa Nova, Natal/RN, CEP 59078-970