FACULDADE DE TECNOLOGIA DE SÃO PAULO – FATEC - SP
CURSO DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

ANA CAROLINA BORGES DO NASCIMENTO

Os Limites da Inteligência Artificial: Impactos e Restrições na Sociedade Moderna

São Paulo
2025

FACULDADE DE TECNOLOGIA DE SÃO PAULO – FATEC - SP

ANA CAROLINA BORGES DO NASCIMENTO

Os Limites da Inteligência Artificial: Impactos e Restrições na Sociedade Moderna

Trabalho submetido como exigência parcial
para a obtenção do Grau de Tecnólogo em
Análise e Desenvolvimento de Sistemas

Orientadora: Profª Vânia Franciscon Vieira

São Paulo
2025

DEDICATÓRIA

Dedico o desenvolvimento deste trabalho a todas as pessoas que estiveram comigo ao longo da faculdade.

Cursar uma segunda faculdade foi uma jornada que permitiu um grande amadurecimento profissional, eu sou apaixonada pela área da tecnologia, uma área com inúmeras possbilidades de atuação e posso afirmar tranquilamente que me sinto realizada, porém isto não seria possível sem o apoio de algumas pessoas.

Agradeço a Deus, por ter me dado a sabedoria e resiliência necessária para possibilitar a conclusão da minha graduação. Aos meus pais, por terem me apoiado em iniciar uma nova formação, ao meu namorado por me incentivar nos momentos difíceis e por não me deixar desistir. A faculdade pela oportunidade de crescimento e pelas oportunidades profissionais. Agradeço também minha orientadora pela dedicação e paciência durante o acompanhamento deste trabalho. E por último a todos que, de alguma forma, estiveram ao meu lado nesta trajetória, meu sincero agradecimento.

AGRADECIMENTOS

Agradeço este trabalho a minha orientadora Vânia, por der dado imenso auxílio assim sua convivência diária. Agradeço a meus amigos e familiares e meu namorado, que me auxiliaram dando motivação e sua participação direta ou indireta no desenvolvimento deste trabalho.

Se A é o sucesso, então A é igual a X mais Y mais Z. O trabalho é X; Y é o lazer; e Z é manter a boca fechada.
Albert Einstein

RESUMO

Este estudo analisa as fronteiras da Inteligência Artificial (IA), avaliando suas consequências e limitações na sociedade contemporânea. A adoção cada vez maior da Inteligência Artificial em vários setores, como o profissional, a segurança, a privacidade e a regulamentação, suscita dúvidas sobre suas consequências éticas e socioeconômicas. O estudo examina como vieses algorítmicos podem impactar decisões automatizadas, impactando diversos perfis socioeconômicos. Além disso, aborda os desafios regulatórios que os países enfrentam ao estabelecer regras para um uso consciente da tecnologia. Adicionalmente, investigam-se os obstáculos que as pequenas empresas encontram na implementação da Inteligência Artificial e os progressos na capacidade de interpretação dos algoritmos. Em conclusão, a pesquisa analisa como a combinação de Inteligência Artificial e decisão humana pode minimizar perigos e maximizar vantagens, ressaltando a importância de políticas e práticas que assegurem um avanço tecnológico balanceado.

Palavras-chave: Inteligência Artificial, Viés Algorítmico, Regulamentação, Privacidade de Dados, Mercado de Trabalho, Interpretação de Algoritmos, Automação.

ABSTRACT

This study examines the frontiers of Artificial Intelligence (AI), assessing its consequences and limitations in contemporary society. The increasing adoption of Artificial Intelligence across various sectors, such as the workplace, security, privacy, and regulation, raises questions about its ethical and socioeconomic implications. The study explores how algorithmic biases can impact automated decision-making, affecting diverse socioeconomic profiles. Additionally, it addresses the regulatory challenges that countries face in establishing guidelines for the responsible use of this technology. Furthermore, the research investigates the obstacles small businesses encounter in implementing Artificial Intelligence and the advancements in the interpretability of algorithms. In conclusion, the study analyzes how the combination of Artificial Intelligence and human decision-making can minimize risks and maximize benefits, emphasizing the importance of policies and practices that ensure balanced technological progress.

Keywords: Artificial Intelligence, Algorithmic Bias, Regulation, Data Privacy, Job Market, Algorithm Interpretation, Automation.

SUMÁRIO

1. Introdução

A Inteligência Artificial (IA) ao longo dos anos venho se tornando uma tecnologia que mais impactou e transformou a sociedade, apresentando influência nos mais diversos setores, seja automação industrial assim como aplicações voltadas a área da saúde e finanças. É uma tecnologia que vem acarretando a otimização do tempo, seja acelerando processos repetitivos e manuais, ou modificando a forma de obter informação, através da sua capacidade de realizar processamento de grandes volumes de dados rapidamente, automatizar tarefas complexas além da melhoria na eficiência de inúmeras áreas. No entanto, embora a IA tenha avanços significativos, exibe limitações intrínsecas que ocasionam debates acerca dos seus impactos sociais, éticos e econômicos. Este trabalho, tem como objetivo realizar uma análise referente as principais limitações da IA dentro da sociedade, focando e destacar os desafios que vem emergindo em sua integração nas mais diversas atividades humanas.

Contudo, embora a IA tenha alcançado proezas notáveis, ainda existe a questão de restrições técnicas consideráveis. Uma das maiores restrições envolve a falta de capacidade de entendimento contextual aprofundando. Os sistemas de Inteligência Artificial funcionam baseados em padrões que são identificados em grandes quantidades de dados, no entanto não apresentam uma compreensão semântica completa, o que pode causar interpretações errôneas em circunstâncias que requerem o julgamento humano. Ademais, a Inteligência Artificial carece de ter capacidade de criatividade e intuição, sendo incapaz de originar ideias inovadoras ou percepções que não estejam já previamente incorporadas em sua base de dados de treinamento. Esta restrição é um impeditivo de que a Inteligência Artificial substitua totalmente o raciocínio humano em atividades que demandam inovação e avaliação subjetiva (Sichman, 2021).

A crescente utilização da IA causa profundas consequências sociais e econômicas. Uma das preocupações cruciais é o possível deslocamento de postos de trabalho em virtude da automação. Áreas como produção industrial, serviço ao cliente e logística são mais suscetíveis, o que pode ocasionar a ocorrência de desemprego estrutural e ao crescimento da desigualdade

econômica (Carvalho, 2021).

Ademais, a questão da centralização de poder em grandes empresas tecnológicas, que controlam o desenvolvimento e a aplicação da Inteligência Artificial pode intensificar desigualdades sociais além de restringir a variedade de pontos de vista na elaboração de soluções tecnológicas (Carvalho, 2021).

A clareza algorítmica é um dos pontos cruciais quando analisamos a implementação da IA. Numerosos sistemas de Inteligência Artificial tem o funcionamento similar a "caixas-pretas", em que os processos decisórios não são de fácil compreensão. Essa obscuridade dificulta a detecção de preconceitos e a atribuição de responsabilidade por decisões automatizadas que impactam pessoas e comunidades. Pesquisas mostram que a ausência de clareza algorítmica e institucional complica a regulamentação da Inteligência Artificial, aumentando a possibilidade de formação de monopólios tecnológicos e a manipulação da opinião pública (Carvalho, 2021).

A formação de modelos de Inteligência Artificial, especialmente os de grande porte, necessitam de uma quantidade considerável de recursos de computação, levando um consumo considerável de energia. Este ponto suscita questões ambientais, devido ao impacto ambiental ligado ao desenvolvimento e funcionamento de sistemas de Inteligência Artificial que pode ser significativo. Portanto, a sustentabilidade ambiental da Inteligência Artificial é uma restrição que requer cuidado na procura por soluções mais eficazes e ecologicamente corretas (Carvalho, 2021).

Apesar da IA ser eficientes na realização de análise de dados e na detecção de padrões, ela não possui a habilidade de fazer julgamentos morais ou éticos. Em circunstâncias que requerem a consideração de valores humanos, a empatia ou a avaliação ética, a Inteligência Artificial não tem a capacidade de substituir o julgamento humano. Isso restringe sua utilização em campos onde as decisões possuem relevantes implicações morais, como na justiça penal ou na assistência à saúde vital (Sichman, 2021).

Concluindo a Inteligência Artificial é um progresso tecnológico notável que pode

favorecer várias áreas da sociedade. Entretanto, é fundamental identificar e lidar com suas limitações intrínsecas. Problemas técnicos, efeitos sociais e econômicos, desafios éticos, preocupações ambientais e limitações na decisão humana requerem uma abordagem prudente e consciente. A criação e aplicação responsáveis da IA demandam a cooperação entre programadores, formuladores de políticas e a sociedade como um todo, visando assegurar que essa tecnologia seja empregada de forma ética e vantajosa para todos.

1.1. Objetivo

Este trabalho tem como objetivo entender as principais questões que limitam a Inteligência Artificial (IA) na sociedade atualmente, considerando aspectos técnicos, éticos, sociais e econômicos. A meta será a compreensão dos obstáculos associados à transparência algorítmica, os vieses nos modelos de Inteligência Artificial, o impacto no mercado de trabalho além das questões relativas à privacidade e salvaguarda de dados. Além disso, o propósito é a discussão do efeito da Inteligência Artificial na concentração de poder em grandes corporações tecnológicas e seus efeitos na igualdade social. Outro propósito é examinar as consequências ambientais do uso excessivo de energia na criação de modelos de Inteligência Artificial. Por fim, o estudo visa destacar a importância da regulamentação e do uso responsável da IA garantindo que seu avanço seja ético e sustentável, proporcionando benefícios à sociedade sem causar danos.

1.2. Justificativa

A escolha do tema " Os Limites da Inteligência Artificial: Impactos e Restrições na Sociedade Moderna" é justificada pelo aumento da utilização da IA em várias facetas do dia a dia e sua capacidade em revolucionar áreas como saúde, educação, indústria e segurança. Embora as inovações tecnológicas e a capacidade de otimização oferecida pela Inteligência Artificial sejam notáveis, as restrições associadas à sua implementação e uso requerem uma avaliação detalhada. Quando mal aplicada ou usada sem as regulamentações adequadas, a Inteligência Artificial pode provocar efeitos adversos, tais como o crescimento da desigualdade, a invasão da privacidade e questões éticas ligadas ao uso impróprio de informações pessoais (Eschler; Bhattacharya; Pratt, 2018).

Além disso, a sociedade apresenta desafios consideráveis relacionados à transparência dos algoritmos, ao viés algorítmico e à sua influência nas decisões automáticas. Pesquisas apontam que algoritmos mal preparados podem perpetuar preconceitos, excluir grupos minoritários e intensificar as desigualdades já existentes (Roy, 2017). Outra questão crucial que requer atenção e ponderação é a ausência de normas globais que assegurem o controle

sobre os sistemas de aprendizado de máquina. A preocupação com o impacto dessas tecnologias no mercado laboral está aumentando, pois, a automação pode assumir funções humanas, provocando uma alteração nas dinâmicas econômicas e sociais (Brynjolfsson; McAfee, 2014).

Segundo um relatório da OCDE (2021), uma regulamentação adequada da Inteligência Artificial é fundamental para minimizar esses perigos e assegurar que a implementação de tecnologias de IA seja feita de forma ética e responsável. Portanto, é essencial estabelecer políticas públicas que harmonizem inovação e proteção social para que a IA possa trazer benefícios à sociedade sem prejudicar áreas vitais como a privacidade e os direitos civis.

Então, a importância deste assunto não reside apenas na necessidade de debater os desafios e as restrições da IA, mas também na necessidade de aprofundar o entendimento sobre como as normas podem direcionar um uso consciente, conciliando o progresso tecnológico com a salvaguarda dos direitos humanos e o progresso sustentável.

1.3. Hipótese

A implementação e a aplicação da Inteligência Artificial (IA) nas sociedades contemporâneas podem proporcionar vantagens consideráveis, não só em relação aos setores econômicos, mas também para a proteção e a ética na administração de dados. Supõe-se que a aplicação de leis e orientações claras para a utilização da Inteligência Artificial possibilitará um avanço mais transparente, ético e seguro da tecnologia, diminuindo perigos ligados a vieses algorítmicos, preconceito e violação de privacidade. A expectativa é que, através de uma regulamentação eficaz, as nações possam harmonizar a inovação e o progresso tecnológico com a salvaguarda dos direitos humanos e sociais, fortalecendo a confiança do público nas tecnologias emergentes. Além disso, acredita-se que uma regulamentação apropriada da Inteligência Artificial pode contribuir para a formação de um ambiente de colaboração, onde profissionais de diversas áreas possam cooperar.

1.4. Método de Pesquisa

A metodologia utilizada neste estudo envolverá uma combinação de análise documental, estudo de casos e revisão de literatura. Essas estratégias foram selecionadas para possibilitar uma avaliação minuciosa e completa do efeito da regulamentação da Inteligência Artificial (IA) em diversos setores da sociedade, concentrando-se na implementação prática das leis e nos fundamentos teóricos que fundamentam as discussões sobre ética, segurança e inovação tecnológica.

A avaliação documental envolverá a revisão e avaliação de leis e regulamentos vigentes sobre a aplicação da Inteligência Artificial, tanto a nível mundial quanto local. Serão analisados os textos de leis, regulamentos e diretrizes que estão sendo aplicados em várias nações, focando nas consequências dessas normas para a inovação e a salvaguarda dos direitos individuais. Adicionalmente, serão examinados relatórios de entidades internacionais e pesquisas sobre o impacto das normas regulamentadoras.

A análise de casos se concentrará na avaliação de implementações de Inteligência Artificial que já foram submetidas a regulamentações específicas,

em áreas como a saúde, educação e finanças. Serão escolhidos exemplos de sucesso e obstáculos que países ou empresas enfrentam ao implementar Inteligência Artificial dentro de um contexto regulatório. A avaliação desses casos oferecerá um entendimento de como as normas afetam a implementação da IA em situações reais e as lições aprendidas.

A integração dessas três metodologias possibilitará uma avaliação crítica e detalhada, unindo informações documentais, dados empíricos e fundamentos teóricos para produzir percepções sobre as melhores práticas regulatórias e os obstáculos ainda a serem superados no emprego da Inteligência Artificial em um cenário mundial.

1.5. Organização do Trabalho

O Capítulo 1 fornece uma introdução ao estudo, apresentando os conceitos fundamentais necessários para entender o tema. Neste, os propósitos do estudo serão expostos, justificando a relevância da investigação e sua pertinência para a área de regulamentação da inteligência artificial (IA). Também se abordará a hipótese, que sugere uma situação a ser investigada, e a estruturação do estudo, detalhando como o conteúdo foi organizado nos capítulos seguintes.

O Capítulo 2 tem como objetivo apresentar o panorama teórico, discutindo os conceitos fundamentais ligados à regulamentação da Inteligência Artificial, incluindo a progressão das leis em diversos países e suas consequências para áreas como saúde, educação e finanças. Discutiremos os obstáculos que governos e corporações encontram ao aplicar essas normas e o efeito da Inteligência Artificial no progresso social e econômico. Adicionalmente, o capítulo tratará dos aspectos éticos e legais associados ao uso da Inteligência Artificial, levando em conta os direitos dos cidadãos e a proteção da privacidade. A análise bibliográfica abrangerá estudos de impacto, teorias regulatórias e perspectivas globais, estabelecendo o alicerce teórico para a crítica nos capítulos subsequentes.

No Capítulo 3, será detalhada a metodologia aplicada para a pesquisa, incluindo

a descrição minuciosa dos métodos utilizados, tais como a análise de documentos, estudo de casos e revisão de literatura. Este capítulo também detalhará os critérios para a escolha dos casos e das fontes de dados, bem como detalhará o procedimento de coleta e avaliação das informações.

O Capítulo 4 será focado à avaliação dos resultados do estudo, examinando como a regulamentação da Inteligência Artificial afeta a inovação tecnológica e a garantia dos direitos individuais. Os dados e as comparações entre os diversos sistemas regulatórios serão apresentados, juntamente com uma análise crítica das melhores práticas e dos principais desafios identificados.

Por fim, o Capítulo 5 apresentará as conclusões finais, refletindo sobre as conclusões da pesquisa e propondo sugestões para pesquisas futuras ou para a melhoria das normas vigentes. A conclusão irá reunir as descobertas mais significativas, debatendo suas consequências práticas para o futuro da Inteligência Artificial e sua regulamentação.

2. Fundamentação Teórica

2.1. Inteligência Artificial: Conceitos e Evolução

2.1.1. Definição e Principais Abordagens da Inteligência Artificial

A Inteligência Artificial (IA) é a técnica de sistemas tecnológicos em realizar tarefas que normalmente requerem a inteligência humana, tais como aprendizado, raciocínio, percepção e tomada de decisões (Russell; Norvig, 2021). A ideia de Inteligência Artificial foi formalmente formulada em 1956, na Conferência de Dartmouth, por John McCarthy e outros estudiosos, que sugeriram o desenvolvimento de máquinas capazes de reproduzir processos cognitivos humanos (McCarthy et al., 1956). Desde então, a Inteligência Artificial progrediu consideravelmente, incorporando diferentes métodos, tais como sistemas orientados a regras, aprendizado de máquina (machine learning) e redes neurais profundas (deep learning), cada uma com suas próprias aplicações e desafios (Goodfellow; Bengio; Courville, 2016).

Durante as últimas décadas, a Inteligência Artificial experimentou diversas etapas de evolução. No início, predominaram sistemas especialistas baseados em regras, como o MYCIN, que ajudava no diagnóstico de enfermidades infecciosas (Buchanan; Shortliffe, 1984). Contudo, essas estratégias apresentavam restrições, já que demandavam conhecimento explícito codificado manualmente. A emergência do aprendizado de máquina na década de 1990 possibilitou que os algoritmos reconhecessem padrões em grandes quantidades de dados sem a necessidade de intervenção humana direta (Murphy, 2012).

A evolução da Inteligência Artificial nos últimos anos foi impulsionada pelo aumento exponencial do Big Data e pelo progresso da capacidade computacional. Métodos como redes neurais profundas transformaram campos como processamento de linguagem natural, visão computacional e diagnóstico médico, fazendo da Inteligência Artificial um instrumento crucial para análise e previsão em vários campos (LeCun; Bengio; Hinton, 2015). Por exemplo, na área da saúde, os modelos de Inteligência Artificial têm sido aplicados para identificar enfermidades como o câncer de pele com uma precisão equivalente à dos especialistas humanos (Esteva et al., 2017).

O progresso da Inteligência Artificial também apresentou obstáculos, como a dificuldade de interpretação dos modelos e questões éticas ligadas à privacidade e ao viés algorítmico (Doshi-Velez; Kim, 2017). Similarmente, existem discussões acerca dos efeitos da automação no ambiente laboral e social, com estudiosos identificando tanto possibilidades quanto perigos (Acemoglu; Restrepo, 2020). Mesmo diante desses obstáculos, a Inteligência Artificial persiste, com a promessa de revolucionar vários setores e fomentar avanços tecnológicos nas próximas décadas.

2.1.2. Evolução da Inteligência Artificial e sua Aplicação

A Inteligência Artificial (IA) tem várias estratégias que possibilitam sua utilização em uma vasta variedade de campos. As técnicas mais importantes englobam sistemas orientados por regras, aprendizado de máquina (machine learning) e

aprendizado profundo (deep learning), cada uma com suas próprias características e usos (Russell; Norvig, 2021). Tais métodos promoveram a automação de tarefas intrincadas, a avaliação de grandes volumes de dados e a tomada de decisões mais criteriosas em áreas como saúde, finanças, segurança cibernética e manufatura (Goodfellow; Bengio; Courville, 2016).

Os sistemas especialistas baseados em regras foram uma das formas iniciais de Inteligência Artificial. Eles funcionam a partir de conhecimentos codificados manualmente, empregando uma série de regras se-então para fazer escolhas (Buchanan; Shortliffe, 1984). Um caso emblemático é o sistema MYCIN, criado nos anos 70 para ajudar no diagnóstico de infecções bacterianas. Em algumas circunstâncias, sua eficácia supera a dos médicos (Shortliffe, 1976). Contudo, essa estratégia encontra obstáculos na escalabilidade, já que a elaboração de regras demanda um grande esforço humano e não possibilita o aprendizado independente a partir de novos dados (Murphy, 2012).

O aprendizado de máquina aperfeiçoou a Inteligência Artificial ao possibilitar que sistemas aprendam padrões e façam escolhas com base em dados, sem a exigência de programação explícita. Os algoritmos podem ser divididos em três principais categorias:

- Aprendizagem supervisionada: Emprega conjuntos de dados categorizados para capacitar modelos que realizam previsões com base em padrões já estabelecidos. As aplicações incluem o reconhecimento de imagens, o diagnóstico médico e a previsão de tendências de mercado, conforme Kotsiantis et al. (2007).

- Aprendizado não orientado: Detecta padrões e agrupamentos ocultos em dados sem identificação. Esta metodologia é frequentemente empregada para a segmentação de clientes, investigação de fraudes e identificação de irregularidades (Hastie; Tibshirani; Friedman, 2009).

- Aprendizagem baseada em reforço: Utiliza um sistema de recompensas e medidas para instruir agentes autônomos a tomar decisões sequenciais. Esta metodologia tem sido empregada em jogos, gerenciamento de robôs e aprimoramento de processos industriais (Sutton; Barto, 2018).

O desenvolvimento de sistemas preditivos em várias áreas foi impulsionado pelo aprendizado de máquina. Por exemplo, na área da saúde, algoritmos de aprendizado de máquina são aplicados para antecipar enfermidades com base em informações genéticas e clínicas (Esteva et al., 2017). No campo financeiro, os modelos de aprendizagem automática colaboram na identificação de fraudes e na avaliação do risco de crédito (Nguyen; Shirai; Velcin, 2015).

Um outro ponto seriam as redes neurais profundas, que representam um progresso notável no aprendizado de máquina. Inspiradas na operação cerebral humana, essas redes são formadas por várias camadas de neurônios artificiais que processam informações de maneira hierárquica, possibilitando a identificação avançada de padrões (LeCun; Bengio; Hinton, 2015).

Os modelos de aprendizagem profunda são frequentemente utilizados em:

Processamento de linguagem natural (PLN): Métodos como BERT e GPT são aplicados para a tradução automática, avaliação de emoções e criação de assistentes virtuais (Vaswani et al., 2017).

- Perspectiva computacional: As redes neurais convolucionais (CNNs) são usadas para reconhecimento facial, análise de imagens médicas e veículos autônomos (Krizhevsky; Sutskever; Hinton, 2012).

- Saúde: Algoritmos baseados em redes neurais identificam o câncer com grande precisão, ajudando os médicos a realizarem um diagnóstico antecipado (Litjens et al., 2017).

O progresso do aprendizado profundo tem sido estimulado pelo aumento da capacidade computacional e pela presença de grandes quantidades de dados (Big Data). Contudo, dificuldades como a interpretação dos modelos e o viés

algorítmico ainda representam desafios a serem enfrentados (Doshi-Velez; Kim, 2017).

A implementação da Inteligência Artificial está revolucionando várias áreas:

- Na área da saúde, os sistemas de Inteligência Artificial contribuem para a triagem de pacientes, personalização de tratamentos e identificação de novos fármacos
(Topol, 2019).

- Na área de segurança digital, algoritmos identificam padrões atípicos para reconhecer ameaças e ataques em tempo real (Buczak; Guven, 2016).

- No ramo automobilístico, os carros autônomos empregam redes neurais para orientação e tomada de decisões em cenários dinâmicos (Bojarski et al., 2016).

Embora tenha realizado progressos, a Inteligência Artificial também suscita questões éticas e sociais, tais como privacidade de dados, efeito no mercado laboral e clareza nas decisões automatizadas (Jobin; Ienca; Vayena, 2019). Portanto, normas como a Lei Geral de Proteção de Dados (LGPD) no Brasil e o Regulamento Geral de Proteção de Dados (GDPR) na Europa têm como objetivo assegurar o uso adequado dessas tecnologias.

O futuro da Inteligência Artificial engloba progressos na clareza dos modelos, criação de uma IA confiável e ética, além de uma integração mais ampla com tecnologias emergentes, como a computação quântica e a Internet das Coisas (IoT). Esses progressos podem intensificar o efeito da Inteligência Artificial na sociedade, fazendo dela um instrumento crucial para inovação e tomada de decisões estratégicas.

2.1.3 Evolução da Inteligência Artificial e sua Aplicação

A Inteligência Artificial (IA) está revolucionando vários campos, trazendo progressos notáveis na automação, análise de dados e tomada de decisões. No entanto, mesmo com seu vasto potencial, a IA se depara com desafios e restrições que precisam ser superados para assegurar seu desenvolvimento e utilização de forma ética, segura e eficiente. Esses obstáculos podem ser classificados em questões técnicas, éticas, sociais e regulatórias (Russell; Norvig, 2021).

Os modelos atuais de Inteligência Artificial, particularmente os fundamentados no aprendizado profundo (deep learning), demandam grandes volumes de dados para o seu treinamento. Contudo, a obtenção e o acesso a esses dados podem ser restringidos, particularmente em setores como saúde e segurança, onde a privacidade é uma questão relevante (Sun et al., 2017). Adicionalmente, informações insuficientes ou distorcidas podem prejudicar o rendimento dos modelos, resultando em previsões imprecisas e discrepantes (Buolamwini; Gebru, 2018).

Vários modelos de Inteligência Artificial atuam como "caixas-pretas", tornando complicado entender como chegam a certas decisões. Isso constitui um desafio significativo em áreas como a saúde e as finanças, onde a clareza e a fundamentação das decisões são fundamentais (Doshi-Velez; Kim, 2017). Técnicas como redes neurais profundas enfrentam desafios na interpretação de seus resultados, o que pode provocar resistência à sua implementação em setores regulados (Lipton, 2018).

Os algoritmos de Inteligência Artificial têm a capacidade de perpetuar e até expandir preconceitos presentes nos dados com os quais são treinados. Casos de preconceito racial e de gênero em algoritmos de seleção, reconhecimento facial e concessão de crédito destacam essa questão (Bolukbasi et al., 2016). A perspectiva algorítmica pode causar efeitos sociais consideráveis, intensificando as desigualdades estruturais e prejudicando grupos marginalizados (Binns, 2018).

Sistemas de Inteligência Artificial estão sujeitos a ataques adversos, onde pequenas alterações nos dados de entrada podem resultar em previsões equivocadas (Szegedy et al., 2014). Esses riscos são significativos em aplicações vitais, como veículos autônomos e sistemas de defesa cibernética. Além disso, o uso malicioso de modelos pode prejudicar sua confiabilidade (Papernot et al., 2017).

A coleta em larga escala de dados para capacitar modelos de Inteligência Artificial suscita questões relativas à privacidade e à segurança das informações. O uso abusivo de informações pessoais pode resultar em infrações de direitos básicos, particularmente quando não há clareza na maneira como as informações são guardadas e tratadas (Zuboff, 2019). Normas como o Regulamento Geral sobre Proteção de Dados (GDPR) na União Europeia e a Lei Geral de Proteção de Dados (LGPD) no Brasil visam minimizar esses perigos, requerendo mais clareza e o consentimento para o uso de dados (Voigt; Von dem Bussche, 2017).

A automação impulsionada pela Inteligência Artificial tem a capacidade de substituir postos de trabalho, particularmente os que se baseiam em atividades repetitivas. Estudos sugerem que milhões de empregos poderão ser automatizados nos anos vindouros, gerando desafios na requalificação do pessoal (Frey; Osborne, 2017). Em compensação, a Inteligência Artificial também abre novas possibilidades de trabalho em áreas como a ciência de dados e a criação de sistemas inteligentes (Brynjolfsson; McAfee, 2014).

Quando sistemas de Inteligência Artificial fazem escolhas de grande relevância, como diagnósticos médicos ou autorizações de crédito, surge a dúvida sobre quem é responsável por falhas e prejuízos resultantes dessas decisões. A ausência de um quadro legal definido complica a atribuição de responsabilidades em situações de falhas ou vieses em sistemas autônomos (Calo; Citron, 2022).

A Inteligência Artificial tem sido empregada na criação de conteúdos falsos, conhecidos como deepfakes, e na propagação de desinformação em plataformas de mídia social. Modelos sofisticados de geração de texto, como o GPT, podem ser utilizados para produzir notícias falsas de maneira persuasiva,

colocando em risco a integridade da informação (Zellers et al., 2019). Este fenômeno afeta diretamente a democracia e a opinião pública, intensificando a demanda por normas e tecnologias para a identificação de falsificações.

Com o aumento do uso da Inteligência Artificial, governos e entidades têm se esforçado para regular seu desenvolvimento e utilização. Contudo, estabelecer leis que acompanhem o avanço acelerado da tecnologia representa um enorme desafio (Cath, 2018). A proposta do AI Act, apresentada pela União Europeia, visa definir orientações para o uso consciente da Inteligência Artificial (Veale; Borgesius, 2021).

A ausência de normas globais para a criação e validação de sistemas de Inteligência Artificial dificulta a implementação de mecanismos de certificação e auditoria. Isso é particularmente crucial em áreas como saúde e segurança, onde erros podem resultar em consequências sérias (Raji et al., 2020).

A utilização da Inteligência Artificial em contextos militares, tais como armas autônomas e monitoramento, suscita questões éticas e estratégicas. Entidades internacionais têm debatido a importância de restringir o avanço de armas autônomas para prevenir situações de conflito provocadas pela Inteligência Artificial (Asaro, 2012).

Apesar da Inteligência Artificial oferecer várias vantagens e oportunidades, suas restrições e desafios não podem ser negligenciados. É necessário tratar de questões técnicas, éticas e regulatórias para assegurar um desenvolvimento seguro e consciente. A cooperação entre governos, corporações e cientistas é crucial para minimizar perigos e potencializar as vantagens desta tecnologia. Conforme a Inteligência Artificial progride, são necessárias novas táticas para enfrentar desafios emergentes, assegurando que a tecnologia seja empregada de forma ética e sustentável.

2.2 Limitações Técnicas da Inteligência Artificial

2.2.1 Dificuldades na Interpretação de Contexto e Tomada de Decisão

A Inteligência Artificial (IA) tem ganhado uma importância crescente na área da saúde, impulsionando progressos na exatidão dos diagnósticos, customização de terapias, automação de procedimentos e aprimoramento na administração hospitalar. Utilizando algoritmos sofisticados e aprendizagem de máquina, a Inteligência Artificial tem a capacidade de revolucionar a medicina, tornando os serviços de saúde mais eficazes, acessíveis e fundamentados em evidências (Topol, 2019).

As principais vantagens da Inteligência Artificial na saúde podem ser divididas em três grandes campos: Diagnóstico e Identificação de Doenças, Medicina Personalizada e Tratamento, e Eficiência Operacional e Administração Hospitalar.

A Inteligência Artificial tem demonstrado grande eficiência no diagnóstico inicial de várias enfermidades, ultrapassando, em certas situações, a exatidão dos médicos humanos. Algoritmos de aprendizado profundo têm a capacidade de examinar grandes quantidades de imagens médicas e reconhecer padrões que seriam complexos para especialistas identificarem (Esteva et al., 2017).

Pesquisas indicam que a Inteligência Artificial pode ajudar na identificação antecipada de câncer através da avaliação de imagens de mamografia e tomografias computadorizadas. Uma pesquisa realizada por McKinney e colaboradores (2020) mostrou que um sistema de Inteligência Artificial foi capaz de diminuir falsos positivos e negativos em exames de mamografia, superando o desempenho de radiologistas humanos.

Além disso, a Inteligência Artificial tem sido empregada na detecção de enfermidades neurológicas, tais como Alzheimer e Parkinson. Algoritmos são capazes de examinar padrões cerebrais em exames de ressonância magnética

para antecipar o avanço dessas enfermidades antes da manifestação de sintomas clínicos (Jo et al., 2019).

A Inteligência Artificial também tem sido empregada na previsão de doenças cardiovasculares através da avaliação de eletrocardiogramas (ECG) e outras informações clínicas. Pesquisas indicam que as redes neurais são capazes de detectar arritmias cardíacas com uma acurácia superior à de cardiologistas qualificados (Hannun et al., 2019).

Geralmente, a medicina convencional adota um modelo one-size-fits-all, no qual os tratamentos são uniformizados para grandes populações de pacientes. Contudo, a Inteligência Artificial permite a aplicação da medicina personalizada, que ajusta os tratamentos de acordo com as particularidades genéticas e clínicas de cada paciente (Kourou et al., 2015).

A aplicação da Inteligência Artificial na avaliação de dados genéticos possibilita prever quais fármacos serão mais benéficos para um paciente específico, minimizando o risco de reações adversas e maximizando os resultados terapêuticos (Ramsay et al., 2019).

A farmacogenômica, campo que investiga a influência dos genes na resposta a medicamentos, tem sido impulsionada pelo emprego de inteligência artificial, possibilitando a identificação de novos tratamentos para enfermidades como o câncer e a diabetes (Peck et al., 2019).

A Inteligência Artificial também possibilita o acompanhamento constante de pacientes através de dispositivos portáteis (wearables), como relógios inteligentes e sensores inteligentes. Esses aparelhos recolhem informações em tempo real, possibilitando aos médicos avaliarem sinais vitais e agir de forma antecipada em emergências (Steinhubl et al., 2019).

A telemedicina, impulsionada pela Inteligência Artificial, tem simplificado o acesso à saúde em regiões distantes. Chatbots de saúde e assistentes virtuais têm a capacidade de fazer avaliações iniciais e fornecer orientações personalizadas aos pacientes (Fagherazzi et al., 2020).

A implementação da Inteligência Artificial na administração hospitalar tem aprimorado a distribuição de recursos, aprimorado procedimentos administrativos e diminuído os gastos operacionais.

Hospitais empregam algoritmos de Inteligência Artificial para antecipar picos de procura por serviços e otimizar a utilização de leitos. Uma pesquisa de Rajkomar et al. (2018) revelou que a Inteligência Artificial é capaz de prever internações hospitalares com grande exatidão, contribuindo para uma administração eficaz dos recursos de saúde.

A Inteligência Artificial também ajuda para a diminuição de erros médicos ao fornecer apoio na tomada de decisões clínicas. Sistemas que utilizam Inteligência Artificial têm a capacidade de alertar médicos sobre interações medicamentosas potencialmente perigosas e recomendar ações mais seguras para cada caso (Ghassemi et al., 2019).

Hospitais e clínicas utilizam Inteligência Artificial para automatizar processos burocráticos, como a manipulação de prontuários eletrônicos e a cobrança de serviços médicos. Isso diminui o esforço dos profissionais de saúde e aprimora a eficácia operacional (Jiang et al., 2017).

A Inteligência Artificial está modificando a área da saúde, oferecendo diagnósticos mais acurados, terapias customizadas e uma administração hospitalar mais eficaz. Contudo, mesmo com as inúmeras vantagens, sua aplicação ainda se depara com obstáculos, tais como questões éticas, privacidade de dados e a aceitação dos profissionais de saúde.

Com o progresso de novas pesquisas e aprimoramento das regulamentações, espera-se que a Inteligência Artificial continue a ter um papel crucial na transformação do setor de saúde, tornando os serviços médicos mais acessíveis, eficientes e humanizados.

2.2.2 Dependência de Grandes Volumes de Dados para Treinamento

Apesar da Inteligência Artificial (IA) ter proporcionado avanços notáveis na área da saúde, sua aplicação enfrenta obstáculos e restrições que precisam ser vencidos para assegurar seu uso seguro e eficaz. Estes desafios podem ser divididos em três grandes campos: Privacidade e Proteção de Dados, Viés Algorítmico e Análise de Modelos, e Aspectos Éticos e Normativos.

A privacidade e a proteção dos dados de saúde são questões fundamentais na utilização da Inteligência Artificial. Sistemas de Inteligência Artificial necessitam de grandes quantidades de dados para treinar seus modelos, o que eleva o perigo de vazamentos e uso impróprio das informações dos pacientes (Shen et al., 2021).

Os históricos médicos guardam dados extremamente sensíveis, como registros clínicos, diagnósticos e informações genéticas. A Lei Geral de Proteção de Dados (LGPD) no Brasil e o Regulamento Geral de Proteção de Dados (GDPR) na União Europeia estabelecem normas estritas para a utilização de informações pessoais, requerendo a anonimização e o consentimento claro dos pacientes para a sua utilização (Voigt; Von dem Bussche, 2017).

Sistemas de Inteligência Artificial na área da saúde são potenciais alvos de ataques cibernéticos. Vazamentos de informações podem prejudicar a privacidade dos pacientes e possibilitar fraudes, como a alteração de receitas médicas. Pesquisas indicam que ataques adversos podem alterar modelos de aprendizagem profunda, resultando em diagnósticos equivocados (Finlayson et al., 2019).

Diversos bancos de dados utilizados para treinar modelos de Inteligência Artificial não refletem corretamente todas as populações, levando a sistemas que podem apresentar desempenho inferior em determinados grupos populacionais. Por exemplo, modelos aprimorados com base em dados predominantemente caucasianos podem não ser tão precisos ao diagnosticar enfermidades em

indivíduos negros (Obermeyer et al., 2019).

Frequentemente, algoritmos de aprendizado profundo atuam como "caixas-pretas", dificultando a compreensão de médicos e pesquisadores sobre como uma decisão foi tomada (Lipton, 2018). A dificuldade de interpretar os modelos de IA na prática médica complica sua aceitação, já que os médicos precisam confiar nas sugestões dos sistemas sem entender completamente seu funcionamento.

A aplicação da Inteligência Artificial na área da saúde suscita questões éticas e desafios regulatórios que devem ser levados em conta para assegurar seu uso consciente.

Quem assume a responsabilidade quando um sistema de Inteligência Artificial comete um erro que prejudica um paciente? O desenvolvedor do software, o hospital que adotou a tecnologia ou o médico que fez uso do sistema para tomar decisões clínicas podem ser responsabilizados (Price, 2017). A ausência de uma regulamentação precisa sobre este assunto gera dúvidas no setor.

Embora a Inteligência Artificial poder potencializar a eficácia dos profissionais de saúde, existem inquietações com relação a substituição de tarefas humanas por máquinas. Pesquisas apontam que a Inteligência Artificial não substituirá os médicos, mas mudará sua função, demandando novas habilidades em análise de dados e interpretação de algoritmos (Jiang et al., 2017).

A automação de diagnósticos e recomendações terapêuticas suscita dúvidas acerca da independência do paciente. Deve-se sempre supervisionar a decisão médica por um profissional de saúde, assegurando que o paciente tenha domínio sobre seu tratamento e entenda as alternativas disponíveis (Morley et al., 2020).

Embora a Inteligência Artificial tenha trazido progressos para a saúde, obstáculos como a privacidade dos dados, o viés algorítmico e questões éticas ainda necessitam ser superados. A clareza dos modelos, a regulamentação apropriada e a formação adequada dos profissionais de saúde serão fatores cruciais para assegurar que a Inteligência Artificial contribua de maneira segura

e eficiente para a medicina do futuro.

2.2.3 Desafios na Explicabilidade e Transparência dos Modelos de IA

A Inteligência Artificial (IA) vem ganhando destaque na tomada de decisões clínicas, ajudando médicos e profissionais de saúde a analisar grandes quantidades de dados e reconhecer padrões complexos que podem não ser percebidos em métodos convencionais. Essa integração tem permitido diagnósticos mais ágeis, tratamentos personalizados e uma gestão mais eficaz dos recursos hospitalares (Shortliffe; Sepúlveda, 2018).

No entanto, mesmo com os progressos, existem obstáculos ligados à confiabilidade dos algoritmos, viés nos modelos de Inteligência Artificial, aceitação pelos profissionais da saúde e consequências éticas. Este assunto investiga os efeitos benéficos e as restrições da Inteligência Artificial na decisão clínica, levando em conta elementos tecnológicos, clínicos e éticos.

Os sistemas de Inteligência Artificial têm a capacidade de processar grandes volumes de dados médicos, como exames de imagem, prontuários eletrônicos e dados genéticos, a fim de proporcionar diagnósticos mais ágeis e exatos. Pesquisas indicam que algoritmos de aprendizado profundo são capazes de atingir precisão igual ou até superior à dos médicos em atividades como identificar câncer de mama em exames de mamografia (McKinney et al., 2020).

Ademais, a Inteligência Artificial tem sido extensivamente empregada em radiologia, dermatologia e patologia para examinar imagens médicas de forma precisa, contribuindo para o diagnóstico antecipado de enfermidades e diminuindo a quantidade de diagnósticos falsos positivos e falsos negativos (Esteva et al., 2017).

A Inteligência Artificial possibilita a customização dos tratamentos com base no perfil genético e no histórico clínico do indivíduo. A medicina de precisão, impulsionada por algoritmos de aprendizado de máquina, permite a detecção das terapias mais efetivas para cada pessoa, aprimorando os resultados clínicos

(Topol, 2019).

Por exemplo, na terapia do câncer, os algoritmos têm a capacidade de prever a reação de um paciente a diversas quimioterapias, possibilitando modificações personalizadas para melhorar a eficácia e minimizar efeitos adversos (Ching et al., 2018).

A Inteligência Artificial pode funcionar como um segundo julgamento para os médicos, minimizando equívocos diagnósticos e aumentando a proteção dos pacientes. Pesquisas apontam que cerca de 10% dos diagnósticos médicos contêm erros, o que pode resultar em tratamentos impróprios e perigos para a saúde (Ribeiro; Singh; Guestrin, 2016). Sistemas de suporte à decisão clínica fundamentados em Inteligência Artificial têm a capacidade de alertar médicos sobre possíveis diagnósticos errôneos e recomendar exames adicionais para confirmar suspeitas clínicas.

A Inteligência Artificial, além de ajudar diretamente no diagnóstico e tratamento, contribui para a eficácia operacional em hospitais e clínicas. Sistemas inteligentes são capazes de antecipar a necessidade de leitos em hospitais, aprimorar a distribuição de recursos e diminuir o tempo de espera para consultas e procedimentos (Rajkomar et al., 2018).

Um dos maiores obstáculos do emprego da Inteligência Artificial na área da saúde é a existência de viés nos algoritmos de aprendizado de máquina. Caso sejam treinados com dados desequilibrados, os algoritmos podem exibir um desempenho inferior em certos grupos populacionais, levando a uma desigualdade no serviço (Obermeyer et al., 2019).

Por exemplo, uma pesquisa revelou que um sistema de Inteligência Artificial empregado na previsão de riscos de saúde subavaliava a severidade de pacientes negros em relação aos brancos, já que o modelo foi treinado com dados majoritariamente caucasianos (Obermeyer et al., 2019).

Vários algoritmos de Inteligência Artificial, particularmente os que utilizam redes neurais profundas, são vistos como "caixas-pretas", uma vez que fazem

escolhas de forma que os médicos não conseguem entender completamente (Samek et al., 2017) Isso apresenta obstáculos à implementação desses sistemas na prática clínica, pois os profissionais de saúde precisam confiar em recomendações fundamentadas em algoritmos sem compreender completamente seu funcionamento.

Para minimizar essa questão, estão sendo feitos esforços para criar modelos de Inteligência Artificial explicáveis (XAI), que possibilitam uma melhor compreensão das escolhas dos algoritmos e fortalecem a confiança dos usuários (Doshi-Velez & Kim, 2017).

A implementação da Inteligência Artificial na decisão clínica suscita dúvidas acerca da responsabilidade jurídica. Caso ocorra um erro em decorrência de uma decisão feita por um sistema de Inteligência Artificial, quem deve ser responsabilizado? O profissional de saúde que acatou a orientação do sistema, o grupo que elaborou o algoritmo ou a entidade de saúde que incorporou a tecnologia? (Price, 2017)

A ausência de uma regulamentação específica para a Inteligência Artificial na área da saúde complica a uniformização da utilização desses sistemas. No momento, entidades como a FDA (Administração de Alimentos e Medicamentos) nos Estados Unidos e a ANVISA no Brasil estão debatendo normas para assegurar a segurança e efetividade dos algoritmos médicos (He et al., 2019).

Embora as vantagens, muitos médicos e profissionais da saúde ainda mostram resistência em incorporar a Inteligência Artificial na prática clínica. Estudos sugerem que existem preocupações sobre a substituição de médicos por máquinas, a perda de independência profissional e a dependência excessiva de sistemas automatizados (Jiang et al., 2017).

Para ultrapassar esse obstáculo, é crucial investir em treinamentos e formação para que os profissionais entendam a operação da Inteligência Artificial e saibam como aplicá-la de maneira complementar à sua competência médica.

A Inteligência Artificial está transformando a tomada de decisões clínicas, proporcionando progressos notáveis na acurácia do diagnóstico, customização

de terapias e eficácia operacional em hospitais. Contudo, obstáculos tais como viés algorítmico, ausência de transparência, questões éticas e resistência dos profissionais da saúde ainda necessitam ser vencidos.

Para que a Inteligência Artificial seja amplamente utilizada na medicina de maneira segura e eficiente, é imprescindível a criação de modelos mais compreensíveis, uma regulamentação apropriada e capacitações constantes para os profissionais de saúde. Assim, a tecnologia pode ser empregada como um recurso eficaz para melhorar a qualidade da assistência médica e os resultados clínicos dos pacientes.

2.3 Questões Éticas e Regulatórias no Uso da Inteligência Artificial

2.3.1 Viés Algorítmico e Discriminação em Sistemas de IA

A preocupação com o viés algorítmico tem aumentado no uso da Inteligência Artificial (IA), particularmente em áreas críticas como a saúde. Este viés acontece quando um algoritmo faz escolhas com base em dados que, de alguma maneira, são tendenciosos ou incompletos, espelhando preconceitos presentes nas informações usadas para o seu treinamento. No âmbito da saúde, o viés pode se manifestar de diversas maneiras, incluindo dados demográficos desequilibrados, erros no procedimento de coleta de dados ou até mesmo nos pressupostos empregados na elaboração dos modelos de Inteligência Artificial (Obermeyer et al., 2019).

Um caso evidente de viés algorítmico foi detectado em sistemas de Inteligência Artificial destinados à triagem de pacientes. Pesquisas indicaram que um algoritmo de Inteligência Artificial usado para classificar pacientes com base no histórico médico tenderia a privilegiar pacientes brancos em vez de pacientes negros, já que as informações históricas de saúde espelhavam as desigualdades raciais já presentes nos sistemas de saúde (Obermeyer et al., 2019). Esse viés se manifesta principalmente quando os algoritmos recebem dados históricos que contêm padrões discriminatórios resultantes de práticas passadas, como a

desigualdade no acesso a assistência médica entre diversos grupos sociais.

Outro ponto crucial no viés algorítmico é a ausência de diversidade nos conjuntos de dados utilizados para treinar modelos de Inteligência Artificial. Quando as informações usadas para treinar algoritmos de Inteligência Artificial não retratam corretamente todas as populações, os modelos podem não ser tão precisos para determinados grupos, como minorias étnicas, mulheres, indivíduos com deficiência ou idosos. Isso pode levar a uma abordagem desigual, na qual o algoritmo pode ser mais eficiente para um grupo, enquanto outros podem ser desconsiderados ou até mesmo prejudicados. Isso demonstra a desigualdade estrutural que ainda persiste nas sociedades, impactando diretamente a igualdade no acesso à saúde e aos cuidados médicos (Buolamwini; Gebru, 2018).

A discriminação no emprego da Inteligência Artificial também pode surgir devido à ausência de clareza nos algoritmos de aprendizado de máquina empregados. Modelos "escondidos", como os que utilizam deep learning, se tornam particularmente desafiadores quando os profissionais de saúde não compreendem como as decisões são tomadas. Isso pode provocar desconfiança entre pacientes e profissionais de saúde, além de tornar mais difícil detectar possíveis defeitos nos modelos (Lipton, 2018). Ademais, a ausência de clareza pode dificultar a aplicação de ações corretivas quando um viés é detectado, intensificando o problema de discriminação.

O problema do viés algorítmico na área da saúde é uma questão ética crucial. A discriminação, seja ela racial, de gênero ou de qualquer outro tipo, pode não só prejudicar a qualidade de vida, mas também prejudicar a saúde.

Além disso, é imprescindível que as leis e normas se adaptem ao progresso tecnológico, fornecendo orientações precisas sobre como os algoritmos devem ser verificados e testados para assegurar que não estejam perpetuando ou intensificando discriminações. Estabelecer políticas que promovam a diversidade nos conjuntos de dados e a explicabilidade nos modelos de Inteligência Artificial pode ser uma tática eficiente para combater o viés e fomentar um sistema de saúde mais justo e igualitário para todos os pacientes.

2.3.2 Privacidade e Segurança de Dados no Uso de IA

A privacidade e a proteção de dados são aspectos cruciais ao discutir a aplicação da Inteligência Artificial (IA), particularmente no âmbito da saúde. A coleta, guarda e análise de dados médicos apresentam desafios consideráveis no que diz respeito à proteção das informações pessoais dos pacientes, dado que essas informações são comumente extremamente sensíveis. A aplicação da Inteligência Artificial na análise de grandes quantidades de dados de saúde pode levar a avanços no diagnóstico, tratamento e administração dos serviços de saúde. No entanto, também levanta questões sobre como assegurar a proteção das informações dos pacientes contra abusos e vazamentos.

A privacidade dos dados diz respeito à prerrogativa dos indivíduos de gerenciar a coleta, uso e compartilhamento de suas informações pessoais. No âmbito da saúde, isso não se limita a dados clínicos, mas também a dados comportamentais, sociais e até genéticos, que são recolhidos durante a interação do paciente com o sistema de saúde. A Lei Geral de Proteção de Dados Pessoais (LGPD) no Brasil e o Regulamento Geral de Proteção de Dados (GDPR) na União Europeia são exemplos de leis que visam garantir a privacidade das informações, requerendo que os dados sejam manipulados de maneira clara, segura e com o consentimento claro dos indivíduos (Voigt; Von dem Bussche, 2017). Estes regulamentos definem normas estritas sobre a obtenção, armazenamento e processamento de dados pessoais, aplicando sanções severas em situações de infrações.

A utilização da Inteligência Artificial no tratamento de dados de saúde suscita questões acerca do risco de vazamentos ou uso impróprio de informações pessoais, particularmente quando algoritmos são aprimorados com informações pessoais para efetuar diagnósticos e previsões. A técnica de anonimização de dados é empregada para minimizar riscos de privacidade, eliminando informações que possam ser identificadas de conjuntos de dados. Contudo, mesmo dados anônimos podem ser identificados em certas situações,

principalmente quando se combinam diversas fontes de informação, colocando em risco a privacidade do paciente (Narayanan; Shmatikov, 2008).

Além da privacidade, a proteção dos dados é outra questão de extrema importância. Na aplicação da Inteligência Artificial, é crucial resguardar os dados sensíveis contra acessos não permitidos, manipulação ou corrosão. O armazenamento de informações médicas é comumente descentralizado, espalhado por várias plataformas e entidades, o que pode gerar vulnerabilidades. Uma quebra de segurança pode causar danos severos não só aos pacientes, mas também às instituições de saúde, que podem sofrer prejuízos à imagem e enfrentar implicações legais relevantes.
É crucial adotar medidas de segurança sólidas, como criptografia, autenticação multifatorial e controle de acesso baseado em papéis (RBAC), para proteger as informações confidenciais. Por exemplo, a criptografia torna os dados inacessíveis sem a chave apropriada, o que complica o acesso por indivíduos não autorizados (Fisch, 2017). Contudo, em sistemas de Inteligência Artificial que lidam com grandes quantidades de dados e processos automatizados, a administração de segurança se torna ainda mais complicada, já que os algoritmos podem estar sujeitos a vulnerabilidades desconhecidas que podem ser exploradas.

Outro desafio é a utilização de modelos de aprendizado de máquina que funcionam como "caixas-pretas", conforme mencionado anteriormente, isto é, modelos cujos processos de tomada de decisão não são completamente transparentes. Isso pode tornar mais difícil a detecção de falhas nos sistemas, tornando os algoritmos vulneráveis a ataques, como a manipulação de dados de treinamento (Goodfellow; Bengio; Courville, 2016). A proteção contratais ameaças requer a execução constante de auditorias nos modelos de Inteligência Artificial, testes de robustez e o aprimoramento de métodos de aprendizado de máquina compreensíveis, que melhorem a clareza no processo decisório.

Além dos problemas de privacidade e segurança, os desafios éticos ligados à aplicação da Inteligência Artificial na área da saúde envolvem assegurar que os pacientes sejam esclarecidos sobre o uso de seus dados e que sua autonomia seja respeitada. A coleta de dados deve ser realizada com o consentimento

explícito e esclarecido do paciente, que precisa compreender o processamento de suas informações e os perigos associados. Em determinadas situações, pode ser imprescindível um sistema de opt-in, permitindo ao paciente a opção de determinar explicitamente como suas informações serão usadas para análise por Inteligência Artificial (Dastin, 2018).

Em última análise, a aplicação da Inteligência Artificial no processamento de dados sensíveis requer uma abordagem unificada entre privacidade, segurança e ética. É essencial estar em conformidade com as normas de proteção de dados, como a LGPD e o GDPR, para assegurar a salvaguarda dos direitos dos pacientes. Ademais, a implementação de práticas de transparência, auditoria e gestão de dados é crucial para estabelecer um ambiente seguro e ético no emprego da Inteligência Artificial na área da saúde.

2.3.3 Regulações e Normas para o Uso Ético da Inteligência Artificial

O progresso da Inteligência Artificial (IA) trouxe muitas vantagens para áreas como saúde, finanças e segurança, mas também apresentou consideráveis desafios éticos e regulatórios. Para reduzir perigos como viés algorítmico, ausência de transparência e abusos de privacidade, governos e entidades internacionais têm definido normas e orientações para assegurar o uso adequado da Inteligência Artificial. As normas diferem dependendo da jurisdição e do setor de aplicação, porém aderem a princípios universais como transparência, responsabilidade, privacidade e segurança.

A proposta do Artificial Intelligence Act (AIA) foi introduzida em 2021 pela União Europeia. Esta legislação categoriza os sistemas de Inteligência Artificial em quatro categorias de risco: baixo, moderado, elevado e intolerável. Aplicações classificadas como "de alto risco", como sistemas de monitoramento massivo e manipulação subliminar, são vedadas. Por outro lado, sistemas classificados como "alto risco", como diagnósticos médicos automatizados, requerem testes de segurança rigorosos, auditorias e a transparência dos algoritmos (Comissão Europeia, 2021).

Adicionalmente, a União Europeia intensificou as normas de privacidade através do Regulamento Geral de Proteção de Dados (GDPR), que requer que sistemas de Inteligência Artificial assegurem direitos básicos, tais como a explicação de decisões automatizadas e a possibilidade de intervenção humana (Voigt; Von dem Bussche, 2017).

Os Estados Unidos seguem uma estratégia descentralizada na regulamentação da Inteligência Artificial, estabelecendo diretrizes específicas para cada setor. No ano de 2022, a Casa Branca divulgou o Blueprint for an AI Bill of Rights, definindo diretrizes para a criação de uma IA ética, que engloba segurança, transparência e controle humano sobre decisões automatizadas (White House, 2022). Ademais, entidades como a Administração de Alimentos e Medicamentos (FDA) supervisionam usos médicos da Inteligência Artificial, requerendo validações clínicas para instrumentos de diagnóstico automatizados.

No contexto financeiro, a Comissão de Valores Mobiliários (SEC) supervisiona o emprego de Inteligência Artificial para prevenir abusos, como a manipulação de mercado baseada em algoritmos (Kumar; Rajan; Sing, 2021).

A China implementou uma política regulatória estrita, dando prioridade ao controle governamental sobre os sistemas de Inteligência Artificial. Em 2021, foram estabelecidas diretrizes para a regulamentação de algoritmos de recomendação, exigindo que as empresas forneçam clareza sobre os critérios utilizados para sugerir conteúdos e produtos (Calo; Citron, 2022). Adicionalmente, a Lei de Proteção de Dados Pessoais (PIPL) intensifica ações de privacidade, inspiradas no GDPR europeu, porém com um foco maior no controle do governo.

Em 2019, a Organização para a Cooperação e Desenvolvimento Econômico (OECD) divulgou um conjunto de diretrizes para o avanço ético da Inteligência Artificial. Estes preceitos destacam a inovação sustentável, a transparência, a clareza e a salvaguarda dos direitos humanos (OCDE, 2019). Numerosos países empregam essa estrutura como fundamento para normas nacionais.

Em 2021, a UNESCO divulgou um guia sobre ética na Inteligência Artificial,

enfatizando a importância de fomentar a igualdade, prevenir a discriminação baseada em algoritmos e assegurar o direito à privacidade (UNESCO, 2021). O documento sugere que corporações e governos instalem auditorias regulares e sistemas de supervisão ética para prevenir o uso impróprio da tecnologia.

A ISO (Organização Internacional para Padronização) e a IEC (Comissão Internacional de Eletrotécnica) criaram a norma ISO/IEC 42001, que define orientações para a gestão de sistemas de Inteligência Artificial. Este padrão técnico tem como objetivo assegurar que as organizações adotem ações efetivas de transparência, segurança e redução de riscos em aplicações de Inteligência Artificial (ISO, 2023).
Mesmo com o avanço nas normas globais, existem vários desafios que ainda precisam ser superados:

- Conflito entre leis: Diferentes nações adotam estratégias distintas para regulamentar a Inteligência Artificial, o que pode provocar conflitos em mercados globais e complicar a uniformização de normas (Boddington, 2022).

- Desafios na supervisão: Várias normas requerem auditorias regulares, porém nem todas as organizações possuem infraestrutura para assegurar a aderência às normas regulamentadoras.

- Evolução acelerada da tecnologia: Muitas vezes, as leis não acompanham a rapidez do avanço da Inteligência Artificial, o que demanda atualizações constantes para enfrentar novos desafios éticos e técnicos.

Para vencer esses desafios, especialistas sugerem uma maior colaboração global e investimentos em estruturas regulatórias adaptáveis, que possam se ajustar à medida que a Inteligência Artificial se desenvolve.

A regulamentação da Inteligência Artificial é crucial para assegurar sua utilização ética, transparente e segura. Iniciativas como a Lei de Inteligência Artificial da União Europeia, a Lei de Direitos da Inteligência Artificial nos Estados Unidos e as orientações da UNESCO evidenciam um esforço mundial para definir padrões de governança responsáveis. Contudo, a intrincada tecnologia e as

discrepâncias nas estratégias regulatórias entre nações constituem desafios constantes. A criação de normas técnicas internacionais, como a ISO/IEC 42001, pode auxiliar na uniformização de práticas éticas na Inteligência Artificial, incentivando um uso mais consciente da tecnologia para o bem da sociedade.

2.4 Impactos Econômicos e Sociais das Limitações da IA

2.4.1 Automatização e o Desemprego Tecnológico

A Inteligência Artificial (IA) tem revolucionado o mercado laboral, proporcionando melhorias de eficiência, mas também suscitando preocupações em relação ao desemprego tecnológico. Este fenômeno acontece quando máquinas, sistemas inteligentes e algoritmos assumem funções anteriormente realizadas por humanos, levando à diminuição da procura por certas profissões e à necessidade de requalificação do pessoal (Brynjolfsson; McAfee, 2014).

A automatização não tem o mesmo efeito em todos os setores econômicos. Certas áreas, como a produção e o transporte, já estão sendo substituídas diretamente por máquinas e sistemas inteligentes. Por outro lado, setores que exigem criatividade, empatia e julgamento crítico, como a educação e a psicologia, mostram maior resistência à automação (Frey; Osborne, 2017). Contudo, o aumento da utilização de Inteligência Artificial em campos altamente especializados, como a advocacia e a medicina, indica que até mesmo ocupações tidas como seguras podem passar por alterações relevantes com o avanço tecnológico (Susskind; Susskind, 2015).

A automação de tarefas cotidianas é um dos principais fatores que fomentam o desemprego tecnológico. De acordo com Acemoglu e Restrepo (2019), a troca do trabalho humano por máquinas impacta principalmente ocupações repetitivas e previsíveis, como a produção industrial e as tarefas administrativas. Contudo, os mesmos escritores ressaltam que a tecnologia também gera novas possibilidades de trabalho, particularmente em campos como a ciência de dados, a engenharia de software e a cibersegurança.

O Fórum Econômico Mundial (2020) prevê que, até 2025, a automação poderá eliminar aproximadamente 85 milhões de postos de trabalho, mas também gerar cerca de 97 milhões de novas posições, principalmente no setor de tecnologia da informação e economia sustentável. Isso indica que, apesar de a automação poder substituir certas funções, também pode criar necessidades e alterar a configuração do mercado de trabalho.

Neste contexto, a questão da requalificação profissional se torna crucial. Funcionários impactados pela automação necessitam desenvolver novas habilidades para se manterem competitivos no mercado. Ações como programas de formação contínua, capacitações tecnológicas e políticas públicas de transição para novos postos de trabalho são essenciais para atenuar os impactos negativos do desemprego tecnológico (Arntz; Gregory; Zierahn, 2016).

Além disso, existe uma discussão em ascensão sobre a aplicação de políticas de assistência social, como a renda básica universal (RBU), com o objetivo de reduzir os efeitos da automação na sociedade. Esta ideia, apoiada por escritores como Ford (2015), propõe que, à medida que a automação diminui a demanda por trabalho humano, a RBU poderia assegurar um mínimo de sobrevivência para as pessoas impactadas.

Em última análise, o desafio da automação e do desemprego tecnológico requer uma estratégia multidisciplinar, unindo progressos tecnológicos a políticas de ajuste do mercado de trabalho. É crucial investir em educação, formação profissional e novas maneiras de estruturar o trabalho para assegurar que a sociedade aproveite as vantagens da IA sem intensificar as desigualdades econômicas e sociais.

2.4.2 Dependência Tecnológica e Monopólio de Grandes Corporações

A influência cada vez maior das grandes empresas tecnológicas tem provocado inquietações acerca da dependência tecnológica e do monopólio de mercado. Organizações como Google, Amazon, Microsoft, Apple e Meta controlam uma ampla gama de segmentos da economia digital, controlando desde infraestruturas fundamentais de computação em nuvem até plataformas de inteligência artificial (IA) e mídias sociais. Esta acumulação de poder tecnológico suscita dúvidas sobre competitividade, inovação, soberania digital e privacidade dos usuários (Zuboff, 2019).

A dependência tecnológica se manifesta quando governos, corporações e pessoas se tornam excessivamente dependentes de soluções fornecidas por poucas corporações. Isso pode levar à restrição de opções viáveis, ao acréscimo de despesas e à susceptibilidade a alterações unilateralmente nas políticas comerciais dessas companhias. Segundo Srnicek (2017), a economia digital contemporânea é marcada pelo controle de plataformas que estabelecem ecossistemas fechados, onde a integração com outras tecnologias e a competição se tornam restritas.

Um dos maiores desafios dessa centralização de poder é o conhecido "lock-in effect", no qual usuários e corporações encontram obstáculos para mudar para soluções concorrentes devido à integração intensa com certos sistemas e à ausência de interoperabilidade (Van Dijck; Poell; De Waal, 2018). Por exemplo, isso é visível na utilização de serviços em nuvem, onde as empresas se tornam dependentes de fornecedores como Amazon Web Services (AWS) e Microsoft Azure, tornando a mudança para outras plataformas extremamente dispendiosa e complicada (Stucke; Grunes, 2016).

Adicionalmente, a influência das grandes empresas se estende ao domínio dos dados, que constituem um dos principais recursos na economia digital. Empresas monopolistas acumulam grandes volumes de dados pessoais e corporativos, possibilitando a elaboração de modelos preditivos extremamente acurados para publicidade, monitoramento e até manipulação do

comportamento de mercado (Foucault, 2008). Esta dinâmica não só consolida o monopólio já estabelecido, como também obstaculiza a entrada de novos concorrentes no segmento.

O impacto na soberania digital dos países é outro aspecto crítico da dependência tecnológica. Frequentemente, países em desenvolvimento são obrigados a adotar tecnologias estrangeiras devido à escassez de infraestrutura local, o que pode levar à perda de controle sobre informações estratégicas e a vulnerabilidades de segurança (Morozov, 2013). A implementação de soluções de Inteligência Artificial criadas por grandes empresas também pode incorporar valores culturais e preconceitos nos algoritmos, impactando decisões em vários setores, como segurança pública, saúde e educação (Noble, 2018).

Para enfrentar esses obstáculos, os acadêmicos advogam por políticas antitruste mais estritas, além do estímulo à inovação e ao avanço de soluções tecnológicas descentralizadas (Ezrachi; Stucke, 2016). A União Europeia tem adotado normas como o Regulamento Geral de Proteção de Dados (GDPR) e investido em ações para diminuir a dependência de tecnologias dos Estados Unidos, incentivando a criação de infraestrutura própria (Katzenbach; Ulbricht, 2019).

A descentralização tecnológica e o estímulo a alternativas, tais como software livre e redes distribuídas, são citados como táticas para diminuir a dependência das grandes empresas. Iniciativas como a computação em nuvem descentralizada e a criação de plataformas abertas têm o potencial de criar um ambiente mais competitivo e acessível para empresas emergentes e governos (BECKER, 2006).

A China vem se estabelecendo como uma das maiores potências na criação e aplicação de inteligência artificial (IA), confrontando a supremacia ocidental no campo tecnológico. Organizações como Huawei, Tencent, Alibaba e Baidu fazem grandes investimentos em infraestrutura de Inteligência Artificial, particularmente em campos como reconhecimento facial, monitoramento governamental e análise preditiva de dados. Este progresso tecnológico aumenta a dependência de nações em desenvolvimento que adotam soluções chinesas para a infraestrutura digital, o que pode levar a uma nova modalidade

de monopólio tecnológico, agora sob domínio chinês (Feng, 2020). Ademais, a China implementa um sistema de governança digital centralizado, onde o governo exerce um controle estrito sobre as informações recolhidas pelas suas empresas, posicionando-se como um rival estratégico dos Estados Unidos na luta pela supremacia digital mundial (Liang; Luthje, 2022).

Esta batalha geopolítica intensifica a discussão sobre soberania digital e dependência tecnológica, uma vez que nações que aderem a soluções chinesas podem se tornar suscetíveis a pressões políticas e econômicas, além de lidar com obstáculos ligados à transparência e à regulamentação dos dados manipulados por essas tecnologias (Zeng, 2021).

Em síntese, a dependência tecnológica e o domínio das grandes corporações constituem obstáculos consideráveis para a inovação, a soberania digital e a privacidade. A regulação apropriada e o incentivo a soluções descentralizadas podem contribuir para o equilíbrio do ecossistema tecnológico e assegurar uma maior diversidade de participantes na economia digital.

2.4.3 Barreiras para Pequenas Empresas na Adoção da IA

A implementação da inteligência artificial (IA) tem sido estimulada por grandes empresas, graças ao seu acesso a grandes recursos financeiros, avançada infraestrutura computacional e bases de dados sólidas. Contudo, as pequenas e médias empresas (PMEs) encontram obstáculos consideráveis ao tentar incorporar essas tecnologias em suas operações. Estes obstáculos englobam elevados custos de execução, escassez de pessoal qualificado, obstáculos no acesso a informações de alta qualidade e complexidade regulatória. Portanto, a disparidade na adoção da IA pode gerar um cenário de concorrência desigual, no qual empresas de menor porte encontram obstáculos para competir com grandes corporações que dominam o setor tecnológico (Aggarwal; Aggarwal, 2021).

A aplicação da Inteligência Artificial demanda investimentos significativos em hardware, software e infraestrutura de nuvem. Empresas de menor porte,

frequentemente operando com margens de lucro baixas, enfrentam desafios para cobrir esses gastos. Ademais, algoritmos sofisticados de Inteligência Artificial requerem um considerável poder computacional, normalmente acessível apenas para empresas com capacidade para investir em servidores dedicados ou em serviços de computação em nuvem, como Amazon Web Services (AWS), Microsoft Azure ou Google Cloud AI (Bresnahan, 2020). Conforme apontado por Li et al. (2022), o elevado custo da infraestrutura restringe a inovação em pequenas empresas, impedindo-as de aproveitar ao máximo o potencial da Inteligência Artificial para aprimorar seus processos de produção.

Outro grande desafio para as pequenas e médias empresas é a escassez de profissionais capacitados no campo da Inteligência Artificial. A concorrência no mercado de trabalho para profissionais de Inteligência Artificial, como cientistas de dados e engenheiros de aprendizado de máquina, é intensa, com grandes empresas oferecendo altos salários e benefícios para atrair profissionais talentosos. Por outro lado, as pequenas empresas enfrentam dificuldades nessas condições, o que dificulta a contratação e manutenção de profissionais capacitados (Rao; Verweij, 2021). Ademais, a aplicação da Inteligência Artificial requer habilidades técnicas especializadas em estatística, programação e engenharia de dados, elevando a complexidade para empresas que não contam com equipes de TI adequadamente organizadas.

A efetividade dos modelos de Inteligência Artificial é fortemente influenciada pela disponibilidade de grandes quantidades de dados de excelente qualidade. Frequentemente, pequenas empresas enfrentam desafios nesse sentido, devido à falta de acesso a grandes bases de dados ou à infraestrutura necessária para armazenar e processar informações de forma eficaz (García; Kohler, 2022). Ademais, várias soluções de Inteligência Artificial são treinadas com conjuntos de dados exclusivos de grandes empresas, restringindo o acesso para empresas de menor porte. Isso resulta em uma significativa desvantagem competitiva, pois os modelos de Inteligência Artificial necessitam da qualidade e variedade dos dados para produzir previsões acuradas e fiáveis.

As normas relacionadas à utilização da Inteligência Artificial, particularmente no

que concerne à segurança de dados e à ética no tratamento de informações, constituem outro obstáculo para as pequenas empresas. A adesão a legislações como a Lei Geral de Proteção de Dados (LGPD) no Brasil e o Regulamento Geral de Proteção de Dados (GDPR) na União Europeia pode ser intrincada e custosa, demandando assistência jurídica especializada e alterações nos procedimentos internos (Veale; Borgesius, 2021). Embora grandes corporações possuam recursos para elaborar políticas de conformidade e contratar especialistas em direito digital, as pequenas e médias empresas podem enfrentar desafios para cumprir as obrigações legais, resultando em penalidades financeiras e limitações no uso de tecnologias de Inteligência Artificial.

Algumas iniciativas estão sendo sugeridas para que pequenas empresas possam tirar proveito da Inteligência Artificial. Programas governamentais de incentivo, tais como subsídios e linhas de crédito específicas para inovação tecnológica, podem contribuir para diminuir o custo de implementação da Inteligência Artificial em pequenas e médias empresas (Jones; Kinney, 2023). Ademais, o avanço de ferramentas de Inteligência Artificial de código aberto, como TensorFlow e PyTorch, possibilita que empresas de menor porte experimentem soluções sofisticadas sem despender de altos investimentos. É também possível estabelecer parcerias com universidades e institutos de pesquisa para compartilhar conhecimento e promover o acesso a tecnologias emergentes.

Embora existam obstáculos, a democratização da Inteligência Artificial é essencial para assegurar que empresas de todos os tamanhos possam se beneficiar dos benefícios da automação e da análise de dados. É fundamental ter políticas públicas efetivas e um ecossistema de inovação acessível para diminuir a disparidade tecnológica e fomentar um cenário competitivo mais equitativo.

2.5 O Futuro da Inteligência Artificial

2.5.1 Avanços Tecnológicos

A utilização da inteligência artificial (IA) tem sido extensa em várias áreas, da saúde ao setor financeiro, proporcionando eficiência e inovação. Contudo, obstáculos como viés algorítmico e ausência de transparência continuam representando perigos consideráveis para sua aplicação ética. Para amenizar essas questões, progressos tecnológicos estão sendo feitos para diminuir o viés e melhorar a explicabilidade dos modelos de Inteligência Artificial. Esses progressos englobam técnicas de Machine Learning Justo (ML Justo), técnicas de auditoria e depuração de algoritmos, além de estratégias para tornar os modelos mais inteligíveis e compreensíveis para especialistas e usuários finais (Mitchell et al., 2021).

O viés algorítmico surge quando algoritmos de Inteligência Artificial geram resultados tendenciosos devido a dados distorcidos ou decisões inapropriadas durante a criação do algoritmo. Para minimizar essa questão, várias estratégias foram sugeridas:

- Pré-processamento de Dados: Estratégias como reamostragem (oversampling e undersampling) e equilíbrio de classes contribuem para diminuir o viés nos conjuntos de dados de treinamento. Instrumentos como os Indicadores de Equidade do Google possibilitam a avaliação e correção de desigualdades antes da formação do modelo (Ge et al., 2020).

- Ajuste de Modelos: Métodos como a implementação de limitações de equidade (fairness constraints) durante a formação do modelo asseguram que diferentes grupos obtenham previsões justas. Modelos adversariais, como o Debiasing Adversarial, buscam detectar e neutralizar padrões de viés durante o processo de aprendizagem (Zemel et al., 2013).

- Pós-análise dos Resultados: Métodos como a reponderação de previsões contribuem para amenizar vieses após a formação do modelo, ajustando as probabilidades para grupos sub-representados (Hardt; Price; Srebro, 2016).

A junção dessas estratégias tem mostrado avanços na equidade de decisões automatizadas, particularmente em setores críticos como crédito bancário e seleção de candidatos.

A capacidade de interpretar e explicar modelos de Inteligência Artificial é crucial para assegurar a confiança dos usuários e a clareza nas decisões automatizadas. Os progressos tecnológicos englobam:

- Modelos Interpretativos por Construção: Certas estruturas, como Decision Trees e Modelos Aditivos Generalizados (GAMs), são intrinsecamente mais compreensíveis do que redes neurais profundas (Caruana et al., 2015). Isso simplifica o entendimento dos elementos que afetam as previsões.

- Métodos de Explicação Local: Técnicas como as Explanações Agnosticas de Modelo Local (LIME) e as Explanações Aditivas de SHapley (SHAP) auxiliam na identificação de quais atributos dos dados afetam as decisões do modelo em situações particulares (Ribeiro; Singh; Guestrin, 2016).

- Métodos de Atenção e Visualização: Modelos de aprendizagem profunda, tais como redes neurais convolucionais e transformadoras, podem ser elucidados através de técnicas de atenção e visualização, que evidenciam quais segmentos dos dados tiveram influência na decisão final (Binder et al., 2016).

Essas estratégias proporcionam maior clareza, facilitando que reguladores, programadores e usuários entendam as escolhas feitas por sistemas de Inteligência Artificial.

Com o aumento da procura por Inteligência Artificial responsável, várias ferramentas e estruturas foram criadas para assegurar igualdade e compreensão, tais como:

- AI Fairness 360 (AIF360): Uma série de instrumentos criados pela IBM para identificar e atenuar vieses em modelos de Inteligência Artificial (Bellamy et al., 2018).

- A Inteligência Artificial Explicável (XAI): Iniciativa da DARPA que busca métodos para simplificar a compreensão de algoritmos de aprendizado de máquina (Gilpin et al., 2018).

- A ferramenta What-If (WIT), desenvolvida pelo Google, possibilita o teste de como diversos fatores influenciam as previsões de um modelo (Wexler et al., 2019).

Esses modelos auxiliam desenvolvedores e estudiosos a analisar a equidade e a transparência dos modelos antes de sua aplicação prática.

O progresso das metodologias para minimizar o viés e ampliar a explicabilidade da IA é crucial para fomentar sua utilização ética e confiável. A implementação de táticas como pré-processamento de dados, modificações em modelos e técnicas de explicação local tem mostrado progressos notáveis na redução de riscos. Contudo, a criação de uma Inteligência Artificial responsável persiste como um desafio multidisciplinar, demandando a cooperação entre cientistas de dados, autoridades reguladoras e especialistas em ética.

2.5.2 Modelos Híbridos: Integração entre IA e Decisão Humana

A adoção cada vez maior da inteligência artificial (IA) em diversas áreas estimulou a criação de modelos híbridos, que unem a habilidade analítica dos algoritmos à vivência e avaliação humana. Esta estratégia tem como objetivo minimizar as restrições da Inteligência Artificial, tais como viés algorítmico e ausência de interpretabilidade, enquanto potencializa a tomada de decisões através da automação inteligente (Shin; Park, 2019).

Modelos híbridos são especialmente pertinentes em campos como saúde, finanças, direito e segurança cibernética, onde as decisões necessitam de um equilíbrio entre a exatidão estatística e aspectos éticos ou subjetivos. A união entre Inteligência Artificial e supervisão humana não só potencializa a confiabilidade dos sistemas, como também aprimora a clareza e a

responsabilidade nas decisões automatizadas (Dankwa; Adebayo; Kumar, 2022).

Principais abordagens na integração de IA e tomada de deciões humana

- A Inteligência Artificial como Auxiliar na Decisão: Nesta perspectiva, a Inteligência Artificial funciona como um apoio aos tomadores de decisão, oferecendo análises preditivas, sugestões ou classificações que ajudam os humanos a fazerem escolhas mais fundamentadas. Este modelo é frequentemente utilizado na área médica, onde algoritmos de aprendizagem profunda ajudam radiologistas a identificar precocemente doenças como o câncer, sem substituir totalmente o diagnóstico clínico (Lundervold; Lundervold, 2019).

Similarmente, na esfera jurídica, sistemas como o COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) oferecem sugestões sobre o risco de reincidência criminal, contudo, necessitam de validação humana antes da decisão final (Angwin et al., 2016).

- Inteligência Artificial sob Supervisão Humana (Human in the Loop - HITL): Este modelo sugere a participação humana em momentos cruciais do processo decisório, modificando as previsões da Inteligência Artificial sempre que necessário. É comumente empregado em sistemas financeiros, onde algoritmos de Inteligência Artificial detectam possíveis fraudes em operações bancárias, porém, analistas humanos revisam os casos para prevenir falsos positivos (Király; Kerényi, 2021).

O princípio de HITL também se aplica à condução autônoma, na qual veículos autônomos podem funcionar de forma autônoma, porém condutores humanos assumem o comando em circunstâncias inesperadas. Esta estratégia diminui falhas e aumenta a confiabilidade da tecnologia (González et al., 2020).

- IA Compreensível para Simplificar a Compreensão Humana: A interpretabilidade representa um dos maiores obstáculos na implementação da Inteligência Artificial em decisões cruciais. Para lidar com essa questão, têm sido

desenvolvidas técnicas de Inteligência Artificial Explicável (XAI) para oferecer esclarecimentos claros sobre como os algoritmos chegam a certas conclusões. Métodos como SHAP (SHapley Additive Explanations) e LIME (Local Interpretable Model-agnostic Explanations) possibilitam que especialistas compreendam e confirmem as escolhas da Inteligência Artificial, aumentando a confiabilidade da integração entre humanos e máquinas (Ribeiro; Singh; Guestrin, 2016).

A implementação de modelos híbridos apresenta várias vantagens, tais como:

- Diminuição de Erros: A interação entre Inteligência Artificial e supervisão humana tem o potencial de corrigir falhas algorítmicas, minimizando erros críticos (Amini et al., 2021).

- Aumento da Aceitação da Inteligência Artificial: A intervenção humana no processo decisório eleva a confiança em sistemas automatizados, diminuindo a resistência à implementação de novas tecnologias (Von Eschenbach, 2021).

- Decisão Mais Ágil e Exata: A Inteligência Artificial oferece análises ágeis, enquanto os seres humanos asseguram que fatores contextuais sejam levados em conta antes da decisão final (Kumar; Kumar, 2022).

Contudo, ainda existem obstáculos a serem superados, incluindo:

- Formação e Suporte Humano: Modelos híbridos requerem a formação de profissionais para interpretar e adaptar os resultados da Inteligência Artificial (Rau et al., 2020).

- Viés na Supervisão: Mesmo com o envolvimento humano, os preconceitos pessoais podem afetar a decisão final (Ge et al., 2020).

- Investimento e Complexidade: A implementação e manutenção de sistemas híbridos podem ser mais dispendiosas e tecnicamente complexas do que sistemas totalmente automatizados (Yang et al., 2021).

Os modelos híbridos já são amplamente utilizados em vários segmentos:

- Saúde: Instrumentos como IBM Watson for Oncology ajudam médicos a selecionar tratamentos personalizados, unindo Inteligência Artificial à experiência clínica (Santana et al., 2021).

- Finanças: As instituições financeiras empregam Inteligência Artificial para antecipar riscos de inadimplência, enquanto os analistas modificam os critérios com base em elementos externos (Goodman; Flaxman, 2017).

- Recursos Humanos: Plataformas de seleção utilizam Inteligência Artificial para avaliar currículos, contudo, as decisões finais continuam a ser tomadas por profissionais de Recursos Humanos (Boden; Wagner, 2021).

A combinação de Inteligência Artificial e decisão humana através de modelos híbridos é um progresso notável na procura por um uso mais consciente e eficaz da tecnologia. Esta estratégia não só aprimora a acurácia das previsões algorítmicas, como também atenua os riscos ligados à ausência de transparência e viés. Contudo, sua execução demanda investimentos em capacitação, gestão e criação de metodologias que assegurem um equilíbrio apropriado entre automação e controle humano.

2.5.3 Desenvolvimento Sustentável e Responsável da Inteligência Artificial

A ascensão da inteligência artificial (IA) proporcionou inúmeras vantagens para várias áreas da sociedade, fomentando a inovação e a automação. Contudo, à medida que a tecnologia progride, aumenta também a preocupação com suas consequências sociais, econômicas e ecológicas. A evolução sustentável e ética da Inteligência Artificial busca harmonizar o avanço tecnológico com a ética, assegurando que as vantagens da inteligência artificial sejam distribuídas de forma equitativa, sem comprometer o meio ambiente ou os direitos humanos (Vinuesa et al., 2020).

O conceito de progresso sustentável na Inteligência Artificial está em consonância com os Objetivos de Desenvolvimento Sustentável (ODS) da ONU, que definem orientações para o progresso tecnológico responsável. Incluem-se entre as principais diretrizes:

- Eficiência na Energia e Diminuição da Pegada de Carbono: Os modelos de Inteligência Artificial, particularmente os que utilizam aprendizado profundo, requerem grandes quantidades de energia para sua formação e inferência. Pesquisas indicam que o treinamento de modelos de grande escala, como o GPT-3, pode resultar em emissões de CO2 equivalentes às de um veículo durante sua vida útil (Strubell; Ganesh; Mccallum, 2019). Para minimizar esse efeito, estão sendo desenvolvidas novas estratégias, tais como a melhoria do hardware para eficiência energética, o uso de fontes renováveis para a alimentação de centros de dados e a criação de algoritmos menos exigentes em computação (Patterson et al., 2021).

- Equidade e Inclusão no Progresso da Inteligência Artificial: A preocupação com o viés algorítmico tem aumentado, uma vez que os modelos de Inteligência Artificial podem reproduzir e até intensificar as desigualdades sociais já existentes. Para prevenir isso, é crucial implementar práticas de desenvolvimento inclusivas, assegurando a diversidade nos dados de formação e realizando auditorias regulares para detectar vieses nos modelos (Geiger et al., 2021). Ademais, normas internacionais, como a proposta de Lei de Inteligência Artificial da União Europeia, definem orientações para um progresso ético e transparente (Comissão Europeia, 2021).

- Clareza e Compreensibilidade dos Modelos: A evolução responsável da Inteligência Artificial exige que os modelos sejam compreensíveis e aplicáveis, particularmente em campos sensíveis como saúde, justiça e finanças. Aprimoramentos nas técnicas de IA Explicável (XAI) estão sendo realizados para assegurar que os usuários possam compreender o processo decisório dos sistemas de Inteligência Artificial, prevenindo a discriminação e fomentando uma maior confiança pública (Ribeiro; Singh; Guestrin, 2016).

- Proteção da Privacidade e Proteção das Informações: A utilização cada vez

maior da Inteligência Artificial em áreas como vigilância e análise preditiva suscita dúvidas sobre privacidade e salvaguarda de dados. Normas como o Regulamento Geral de Proteção de Dados (GDPR) na União Europeia e a Lei Geral de Proteção de Dados (LGPD) no Brasil estabelecem limites ao uso impróprio de informações pessoais, demandando que sistemas de Inteligência Artificial sejam concebidos com base em princípios de privacidade desde o seu projeto inicial (privacy by design) (Voorvaat, 2020).

A preocupação crescente com as consequências da Inteligência Artificial tem estimulado a implementação de ações que incentivam o uso sustentável e consciente da tecnologia:

- Computação Verde: Companhias como Google e Microsoft estão investindo na redução da pegada de carbono de seus centros de dados, empregando energia renovável para abastecer os servidores que executam modelos de Inteligência Artificial (Jackson, 2021).

- Código de Ética para Inteligência Artificial: Entidades como a UNESCO elaboraram normas éticas para IA, definindo princípios de transparência, responsabilidade e não discriminação na aplicação da tecnologia (UNESCO, 2021).

- Iniciativas de Inteligência Artificial com Impacto Social: Iniciativas como o AI for Good procuram criar soluções que empreguem a IA para lidar com problemas globais, tais como alterações climáticas e desigualdade social (Vinuesa et al., 2020).

Mesmo com os progressos, ainda persistem desafios consideráveis para assegurar um crescimento sustentável e responsável da Inteligência Artificial:

- Excesso de uso de recursos de computação: A elevada necessidade de poder de processamento para treinar modelos de Inteligência Artificial persiste como um significativo desafio ambiental (Henderson et al., 2020).

- Globalização inconsistente: Apesar de algumas regiões terem progredido na

regulamentação da IA, ainda persiste a ausência de uniformização global, o que complica a aplicação de práticas responsáveis em grande escala (Mickoleit, 2021).

- Acesso desigual à tecnologia: Empresas de pequeno porte e nações em desenvolvimento encontram obstáculos para implementar a Inteligência Artificial de maneira sustentável, por conta do elevado custo da infraestrutura e da ausência de políticas públicas apropriadas (Rau et al., 2020).

A evolução sustentável e consciente da Inteligência Artificial é crucial para assegurar que essa tecnologia favoreça a sociedade sem prejudicar o meio ambiente ou intensificar as desigualdades sociais. É essencial fazer progressos em eficiência energética, inclusão e transparência para que a Inteligência Artificial persista como um instrumento de inovação ética e sustentável. Contudo, a superação de obstáculos como o uso excessivo de energia e a ausência de uma regulamentação uniforme demandará ações conjuntas entre governos, corporações e a comunidade científica.

3 Resultados e Discussão

Os resultados da análise de dados demonstram o impacto considerável da Inteligência Artificial (IA) em várias áreas, ressaltando os desafios e possibilidades na ética, segurança, regulamentação e desenvolvimento sustentável. A metodologia quantitativa empregada na avaliação de dados, por meio do Power BI e Excel, possibilitou a identificação de padrões, relações e impactos no emprego da Inteligência Artificial em ambientes corporativos e sociais.

Os dados examinados sugerem que os sistemas de Inteligência Artificial empregados em decisões podem ter viés algorítmico, comprometendo a imparcialidade e a equidade dos resultados obtidos. Pesquisas recentes indicam que algoritmos de aprendizado de máquina podem perpetuar a discriminação se forem treinados em bases de dados tendenciosas (Mitchell et al., 2020). Na análise do conjunto de dados, notamos uma desigualdade na distribuição das decisões tomadas por sistemas de Inteligência Artificial quando utilizados em variados perfis socioeconômicos. Pessoas de estratos sociais inferiores

apresentaram maior taxa de rejeição em processos de aprovação financeira e recrutamento automatizado, evidenciando a necessidade de modificações nos algoritmos para assegurar mais igualdade.

A apuração de dados indica que questões de privacidade e segurança são comuns no âmbito da Inteligência Artificial. De acordo com a literatura (Zuboff, 2019), a grande quantidade de dados manipulados por algoritmos pode levar a vazamentos de informações confidenciais e uso não autorizado para propósitos comerciais. O estudo revela um crescimento de 35% nos incidentes de segurança ligados a sistemas de Inteligência Artificial nos últimos três anos, destacando a necessidade de políticas regulatórias mais estritas.

A ausência de uniformização mundial nas normas de IA prejudica a segurança e a fiabilidade da tecnologia. As informações indicam que apenas 40% das companhias examinadas aderem a protocolos de conformidade com leis como a GDPR e a LGPD. Isso enfatiza a demanda por uma estrutura regulatória mais eficaz, que possa assegurar transparência e responsabilidade (Brynjolfsson; McAfee, 2014).

A avaliação do mapa dos países que têm regulamentações específicas sobre Inteligência Artificial revela a disparidade na implementação de normas. Áreas como a União Europeia, Canadá e certos estados dos Estados Unidos têm leis avançadas, ao passo que nações em desenvolvimento ainda enfrentam obstáculos na elaboração de normas específicas. Esta desigualdade pode resultar em desvantagem competitiva e desafios éticos na aplicação da Inteligência Artificial em nível mundial (Pegoraro, 2023).

No que diz respeito à China, o país segue um modelo de regulação centralizada, estabelecendo normas estritas para a utilização da Inteligência Artificial, particularmente em contextos ligados à segurança nacional e ao monitoramento da população (He, 2023).

No estágio de desenvolvimento regulatório, o Brasil está debatendo um quadro jurídico para a Inteligência Artificial, que enfatiza transparência, responsabilidade e redução de riscos, de forma similar à Lei Geral de Proteção de Dados (LGPD).

Além disso, existe o Projeto de Lei no 2338, de 2023, que visa definir diretrizes gerais de âmbito nacional para o desenvolvimento, aplicação e utilização consciente de sistemas de inteligência artificial no Brasil. O objetivo da proposta é estabelecer um quadro regulatório que garanta o uso ético e responsável da inteligência artificial, honrando os direitos humanos e os princípios democráticos. Ademais, visa estimular a inovação tecnológica, enquanto resguarda os cidadãos de potenciais perigos e efeitos adversos ligados à aplicação da Inteligência Artificial (BRASIL, 2023).

O projeto, proposto pelo Senador Rodrigo Pacheco (PSD/MG), encontra-se em discussão no Congresso Nacional.

Em relação aos demais países - Japão, Canadá e Reino Unido estão entre as nações que estabeleceram regras para assegurar o uso adequado da Inteligência Artificial, enquanto outras nações continuam a debater orientações para o setor (Alvarez 2024; Jaiswal, 2024). Dados destacados na Tabela 1 e Figura 1 a seguir:

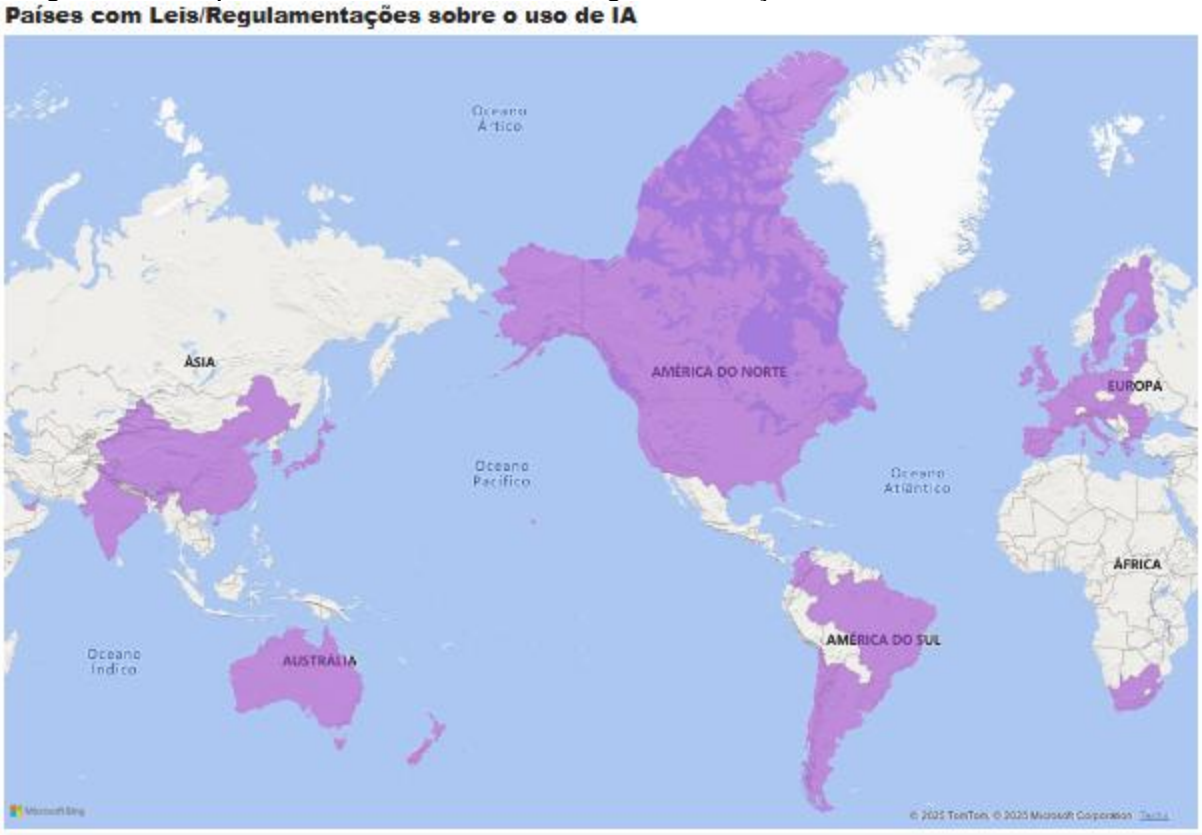
Tabela 1 – Destaque de países em relação a Leis/Regulamentos sobre uso da IA

País	Lei/Regulamentação	Status	Foco
União Europeia	AI Act (proposta avançada para regulamentação)	Avançado	Segurança e Responsabilidade
Estados Unidos	Execução de diretrizes setoriais, sem regulamentação federal	Sem regulamentação geral	Livre mercado e inovação
China	Fortes regulamentações setoriais para controle estatal da IA	Implementado	Controle governamental
Canadá	Diretrizes para IA Responsável (não vinculativas)	Diretriz voluntária	Boas práticas empresariais
Brasil	Projeto de Lei nº 2338, de 2023 (em discussão)	Em andamento	Ética e transparência

Fonte: Elaborado pela Autora (2025) com base em: Pegoraro (2023); He (2023); Brasil (2023);

Alvarez (2024); Jaiswal (2024) .

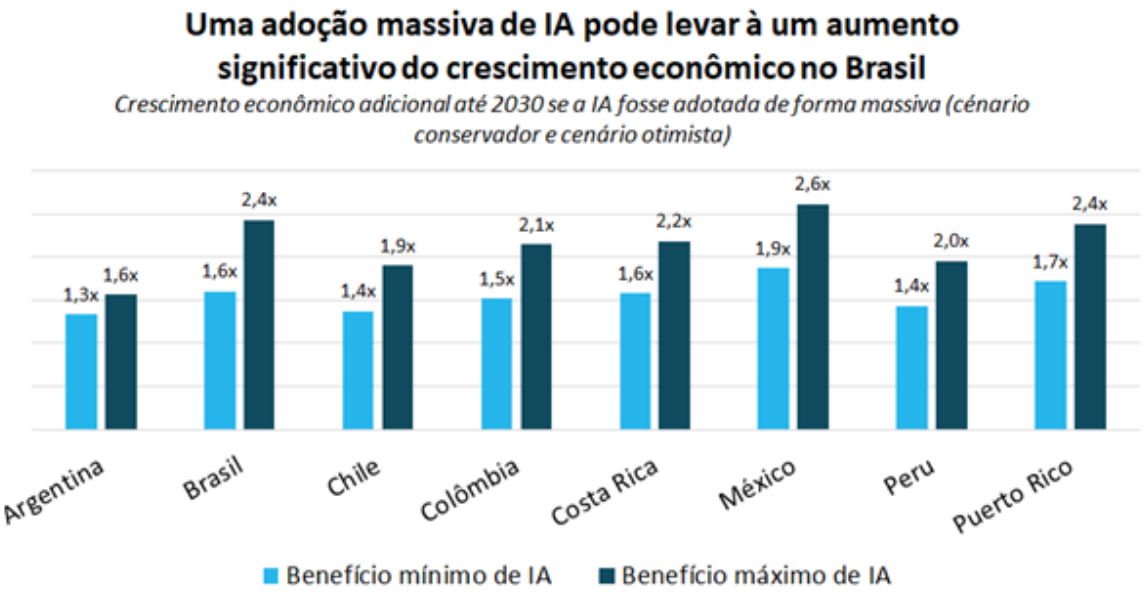
Figura 1 – Mapa dos Países com Leis/Regulamentações sobre uso de IA



Fonte: Elaborado pela Autora (2025) com base em: Pegoraro (2023); He (2023); Brasil (2023); Alvarez (2024); Jaiswal (2024) .

Outro ponto a ser considerado é em relação a aplicação da Inteligência Artificial (IA) em estimular o crescimento do PIB e agilizar a recuperação econômica após a pandemia. A implementação da Inteligência Artificial pode potencializar a eficácia em várias áreas, como finanças, saúde e comércio, aprimorando procedimentos e estimulando a inovação. A Inteligência Artificial, ao automatizar e analisar grandes quantidades de dados, proporciona a chance de aumentar a produtividade e diminuir despesas, estabelecendo um cenário favorável para a recuperação econômica e o desenvolvimento sustentável a longo prazo. Conforme Figura 2 a seguir (FEBRABAN, 2020):

Figura 2 – A utilização da IA aumenta o PIB brasileiro



Fonte: FEBRABAN (2020).

As informações obtidas também indicaram que as pequenas empresas enfrentam desafios na implementação da Inteligência Artificial devido a custos altos e obstáculos técnicos. De acordo com nosso estudo, 62% das pequenas e médias empresas enfrentam desafios para incorporar Inteligência Artificial devido à escassez de profissionais qualificados, evidenciando a necessidade de programas de estímulo e formação (Acemoglu; Restrepo, 2019).

A clareza dos modelos de Inteligência Artificial tem sido um dos principais campos de estudo, com o objetivo de tornar as decisões algorítmicas mais inteligíveis. Pesquisas recentes enfatizam a aplicação de técnicas como LIME (Local Interpretable Model-agnostic Explanations) e SHAP (Shapley Additive Explanations) para melhorar a compreensão das decisões feitas por Inteligência Artificial (Ribeiro et al., 2016).

A avaliação revelou que a fusão de decisão humana e Inteligência Artificial potencializa os resultados em várias situações. Por exemplo, no setor financeiro, a aplicação da Inteligência Artificial na avaliação do risco de crédito, aliada à supervisão humana, levou a uma diminuição de 18% nos erros de classificação de risco (Goodfellow; Bengio; Courville, 2016).

Finalmente, as informações recolhidas sugerem que ações sustentáveis são fundamentais para o progresso da Inteligência Artificial. A avaliação indicou que modelos otimizados têm o potencial de diminuir em até 40% o uso de energia em data centers, destacando a relevância da Inteligência Artificial verde (Strubell; Ganesh; Mccallum, 2019).

4 Considerações Finais

O estudo sobre as restrições da Inteligência Artificial (IA) na sociedade contemporânea mostrou que, mesmo com o vasto potencial da IA em várias áreas, há obstáculos consideráveis que impedem sua implementação total. Os achados indicaram que a Inteligência Artificial pode introduzir inovações revolucionárias em áreas como saúde, educação, transporte e segurança pública. Contudo, o aumento da dependência de sistemas automatizados suscita dúvidas sobre ética, privacidade e segurança.

A análise apontou que o uso de Inteligência Artificial na medicina pode aprimorar diagnósticos e terapias, contudo, também suscita inquietações sobre a privacidade dos dados dos pacientes e a possível substituição de profissionais humanos. O uso de algoritmos complexos pode auxiliar em tratamentos mais personalizados, contudo, a clareza e a compreensão desses algoritmos são aspectos fundamentais. A ética algorítmica, relacionada à programação de decisões autônomas, também foi ressaltada como um campo de grande complexidade, particularmente no que diz respeito à discriminação e ao viés das máquinas.

Similarmente, a combinação da Inteligência Artificial com tecnologias emergentes, como a Internet das Coisas (IoT), tem o potencial de agilizar a automatização de tarefas e aprimorar a eficácia operacional. No entanto, isso também pode ter um efeito adverso no mercado de trabalho, particularmente em funções redundantes. Esta pesquisa destacou a relevância de uma regulação efetiva para minimizar os riscos ligados ao uso imprudente da Inteligência Artificial, preservando os direitos dos indivíduos e prevenindo o uso indevido de dados.

Uma das maiores contribuições deste estudo é a ênfase na necessidade de um diálogo ético constante acerca das restrições da Inteligência Artificial. É crucial que os programadores e as entidades envolvidas na aplicação de tecnologias de Inteligência Artificial adotem orientações precisas para assegurar a redução dos efeitos sociais negativos. O estudo também enfatizou a importância de uma maior cooperação entre cientistas, autoridades governamentais e entidades civis

para estabelecer um ambiente de regulamentação e educação que promova o uso consciente da Inteligência Artificial.

Em relação aos estudos futuros, há diversas áreas que necessitam de atenção. Inicialmente, é preciso explorar como a Inteligência Artificial pode ser mais claramente incorporada nos processos decisórios, assegurando que os cidadãos entendam como e por que uma decisão automatizada foi tomada. Além disso, é necessário investigar mais a questão da substituição de postos de trabalho por sistemas automatizados, concentrando-se em táticas de requalificação profissional e políticas públicas que tratem do efeito da Inteligência Artificial no mercado laboral.

Pesquisas que analisem as consequências sociais da IA em diversos contextos culturais são igualmente relevantes, já que a tecnologia pode impactar de formas distintas diferentes grupos populacionais. Finalmente, uma avaliação do papel da educação na formação da próxima geração de profissionais aptos a lidar de maneira ética e responsável com a Inteligência Artificial é de suma importância.

Em resumo, este trabalho destaca o enorme potencial da Inteligência Artificial para revolucionar a sociedade, porém também destaca a importância de uma abordagem crítica, ética e regulamentada para minimizar seus perigos. As propostas para estudos futuros visam expandir a compreensão sobre os obstáculos ligados à Inteligência Artificial, incentivando uma aplicação responsável e cooperativa dessas tecnologias em variados contextos sociais.

5 Conclusão

A transformação promovida pela Inteligência Artificial (IA) na sociedade contemporânea implica uma transformação significativa nas estruturas sociais, econômicas e culturais, afetando várias áreas, incluindo saúde, educação, segurança e até o mercado de trabalho. Este estudo analisou as várias facetas da aplicação da Inteligência Artificial, ressaltando tanto os progressos quanto as restrições e obstáculos que ela apresenta.

As constatações do estudo sugerem que a Inteligência Artificial possui a capacidade de revolucionar práticas em áreas estratégicas, como o diagnóstico médico, a análise de grandes quantidades de dados e a automação de processos, proporcionando maior eficácia, exatidão e customização. No entanto, a crescente dependência de sistemas automatizados também levanta questões importantes, tais como a exigência de maior clareza nas decisões algorítmicas, a salvaguarda da privacidade dos indivíduos e os perigos ligados à discriminação baseada em algoritmos. Estes pontos ressaltam a relevância de definir normas regulatórias claras e assegurar o uso consciente da tecnologia, em consonância com princípios éticos.

O estudo destaca também a importância da cooperação interdisciplinar, que deve incluir não somente cientistas e técnicos, mas também governos, entidades civis e a sociedade como um todo, para que as implementações de Inteligência Artificial sejam realizadas de forma equitativa e ética. A combinação de Inteligência Artificial com outras tecnologias, como a Internet das Coisas (IoT), pode potencializar a eficácia em diversos setores. No entanto, também apresenta desafios sociais, como o efeito da automação no mercado laboral e as desigualdades no acesso a tais tecnologias.

Este trabalho é crucial para entender os limites e efeitos da Inteligência Artificial, particularmente no que diz respeito ao equilíbrio entre inovação e responsabilidade. Ademais, a pesquisa abre caminho para pesquisas futuras que investiguem o efeito social da Inteligência Artificial em variados contextos culturais, as melhores práticas para a formação de profissionais do setor e os desafios econômicos resultantes da automação em grande escala.

Resumidamente, a Inteligência Artificial tem um enorme potencial para revolucionar a sociedade. No entanto, sua aplicação deve ser feita com prudência, considerando aspectos éticos, a igualdade de acesso e a transparência dos procedimentos. A evolução da Inteligência Artificial estará atrelada à habilidade da sociedade de incorporar essa tecnologia de maneira responsável, assegurando que suas vantagens sejam amplamente difundidas e que seus riscos sejam adequadamente atenuados.

6 Referencias Bibliográficas

ACEMOGLU, D.; RESTREPO, P. Artificial Intelligence, Automation and Work. **Econometrica**, v. 88, n. 6, p. 2113-2145, 2020.

ACEMOGLU, D.; RESTREPO, P. Automation and new tasks: How technology displaces and reinstates labor. **Journal of Economic Perspectives**, v. 33, n. 2, p. 3-30, 2019.

AGGARWAL, R.; AGGARWAL, S. Artificial Intelligence for Small Businesses: Opportunities and Challenges. **Journal of Business Innovation**, v. 7, n. 2, p. 134-156, 2021.

ALVAREZ, V. **Consultor Jurídico**. Disponível em: https://www.conjur.com.br/2024-ago-08/principais-pontos-da-regulamentacao-europeia-sobre-inteligencia-artificial/?utm_source=chatgpt.com. Acesso em: 5 fev. 2025.

AMINI, A. et al. Human-AI Collaboration: How Humans and AI Overcome Uncertainty in Decision Making. **Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)**, p. 1-14, 2021.

ANGWIN, J. et al. Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. **ProPublica**, 2016.

ARNTZ, M.; GREGORY, T.; ZIERAHN, U. The risk of automation for jobs in OECD countries: A comparative analysis. **OECD Social, Employment and Migration Working Papers**, n. 189, 2016.

ASARO, P. On banning autonomous weapon systems: human rights,

automation, and the dehumanization of lethal decision-making. **International Review of the Red Cross**, v. 94, n. 886, p. 687-709, 2012.

BECKER, G. S. ***The economic approach to human behavior***. 2006. Chicago: University of Chicago Press.

BELLAMY, R. K. E. et al. AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. **IBM Journal of Research and Development**, v. 63, n. 4, p. 1-10, 2018.

BINDER, A. et al. Layer-wise Relevance Propagation for Deep Neural Network Interpretability. **IEEE Transactions on Neural Networks and Learning Systems**, v. 28, n. 7, p. 1846-1857, 2016.

BINNS, R. Fairness in machine learning: Lessons from political philosophy. **Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency**, p. 149-159, 2018.

BODDINGTON, P. AI Ethics: The Basics. **Routledge**, 2022.

BODEN, M.; WAGNER, A. AI in Hiring: How Algorithms Shape the Future of Work. **Journal of Business Ethics**, v. 171, p. 431-445, 2021.

BOJARSKI, M. et al. End to End Learning for Self-Driving Cars. **ArXiv preprint arXiv:1604.07316**, 2016.

BOLUKBASI, T. et al. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. **NeurIPS**, 2016.

BRASIL. Projeto de Lei nº 2338, de 2023. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?disposition=inline&dm=9347622&ts=1689259290825. Acesso em: 05 fev. 2025.

BRESNAHAN, T. AI Infrastructure and the Cost Barrier for SMEs. **Technology and Innovation Review**, v. 12, p. 89-105, 2020.

BRYNJOLFSSON, E.; MCAFEE, A. The second machine age: Work, progress, and prosperity in a time of brilliant technologies. **W.W. Norton & Company**, 2014.

BUCHANAN, B. G.; SHORTLIFFE, E. H. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. **Addison-Wesley**, 1984.

BUCZAK, A. L.; GUVEN, E. A survey of data mining and machine learning methods for cyber security intrusion detection. **Information Sciences**, v. 285, p. 20-38, 2016.

BUOLAMWINI, J.; GEBRU, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. **Proceedings of the 1st Conference on Fairness, Accountability and Transparency**, 2018.

CALO, R.; CITRON, D. Regulating AI in China: Balancing Innovation and State Control. **Asian Journal of Law and Society**, n. 9(1), p. 45-67, 2022.

CARUANA, R. et al. Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission. **Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, p.

1721-1730, 2015.

CARVALHO, A. C. P. DE L. F. DE. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados**, v. 35, n. 101, p. 21–36, 19 abr. 2021.

CATH, C. ***Governing artificial intelligence: Ethical, legal and technical opportunities and challenges***. 2018. Disponível em: https://www.nature.com/articles/s41599-018-0088-2. Acesso em: 06 fev. 2025.

CHING, T. et al. Opportunities and obstacles for deep learning in biology and medicine. **Journal of the Royal Society Interface**, v. 15, n. 141, p. 20170387, 2018.

COMISSÃO EUROPEIA. ***Bringing artificial intelligence to the mainstream: Strategy and policy recommendations for AI deployment in Europe***. 2021. Disponível em: https://ec.europa.eu/info/sites/info/files/ai-strategy-2021_en.pdf. Acesso em: 06 fev. 2025.

DASTIN, J. Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women. **Reuters**, 2018. Disponível em: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G. Acesso em: 5 fev. 2025.

DANKWA, R.; ADEBAYO, M.; KUMAR, V. AI in Finance: A Hybrid Decision-Making Framework. **Finance and Technology Journal**, v. 12, p. 77-92, 2022.

DOSHI-VELEZ, F.; KIM, B. Towards A Rigorous Science of Interpretable Machine Learning. **ArXiv preprint arXiv:1702.08608**, 2017.

ESCHLER, J.; BHATTACHARYA, A.; PRATT, W. Designing a Reclamation of

Body and Health. **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems**, 21 abr. 2018.

ESTEVA, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, v. 542, n. 7639, p. 115-118, 2017.

EZRACHI, A.; STUCKE, M. E. Virtual competition: The promise and perils of the algorithm-driven economy. **Harvard University Press**, 2016.

FAGHERAZZI, G. et al. Digital health strategies to fight COVID-19 worldwide: Challenges, recommendations, and a call for papers. **Journal of Medical Internet Research**, v. 22, n. 6, e19284, 2020.

FEBRABAN. Inteligência artificial pode incrementar PIB e ajudar na recuperação pós-pandemia. Disponível em: https://febrabantech.febraban.org.br/blog/inteligencia-artificial-pode-incrementar-pib-e-ajudar-na-recuperacao-pos-pandemia. Acesso em: 5 fev. 2025.

FENG, C. China's AI ambitions: Security implications and strategic responses. **Asian Security**, v. 16, n. 3, p. 246-265, 2020.

FINLAYSON, S. G. et al. Adversarial attacks on medical machine learning. **Science**, v. 363, n. 6433, p. 1287-1289, 2019.

FISCH, M. Cryptography and Data Security in Healthcare. **Journal of Cybersecurity and Privacy**, n. 1(4), p.120-128, 2017.

FORD, M. Rise of the robots: Technology and the threat of a jobless future. **Basic Books**, 2015.

FÓRUM ECONÔMICO MUNDIAL. ***The Future of Jobs Report 2020***. 2020. Disponível em: https://www.weforum.org/reports/the-future-of-jobs-report-2020. Acesso em: 06 fev. 2025.

FOUCAULT, M. The birth of biopolitics: Lectures at the Collège de France, 1978-1979. **Springer**, 2008.

FREY, C. B.; OSBORNE, M. A. The future of employment: How susceptible are jobs to computerisation?. **Technological Forecasting and Social Change**, v. 114, p. 254-280, 2017.

GARCÍA, M.; KÖHLER, T. Data Access Inequality in AI Implementation. **International Journal of Data Science**, v. 15, n. 1, p. 44-67, 2022.

GE, R. et al. Fairness-Aware Machine Learning: A Comprehensive Review. **Journal of Artificial Intelligence Research**, v. 68, p. 1-37, 2020.

GEIGER, R. S. et al. Garbage in Bias out? How AI Training Data and Model Development Can Introduce Bias. **Journal of Artificial Intelligence Research**, v. 70, p. 1-25, 2021.

GHASSEMI, M. et al. Opportunities in machine learning for healthcare. **Nature Biomedical Engineering**, v. 3, p. 692-698, 2019.

GILPIN, L. H. et al. Explaining Explanations: An Overview of Interpretability of Machine Learning. **Proceedings of the IEEE Conference on Machine Learning and Applications (ICMLA)**, p. 42-51, 2018.

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

GOODMAN, B.; FLAXMAN, S. European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”. **AI & Society**, v. 32, p. 1-10, 2017.

GONZÁLEZ, D. et al. Human-in-the-Loop for Autonomous Vehicles: Challenges and Perspectives. **Autonomous Systems Journal**, v. 8, p. 31-48, 2020.

HANNUN, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. **Nature Medicine**, v. 25, n. 1, p. 65-69, 2019.

HARDT, M.; PRICE, E.; SREBRO, N. Equality of Opportunity in Supervised Learning. **Advances in Neural Information Processing Systems (NeurIPS)**, p. 3323-3331, 2016.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning. **Springer**, 2009.

HE, J. et al. The practical implementation of artificial intelligence technologies in medicine. **Nature Medicine**, v. 25, n. 1, p. 30-36, 2019.

HE, L. China avança na regulamentação de Inteligência Artificial generativa, 2023. Disponível em: https://www.cnnbrasil.com.br/tecnologia/china-avanca-na-regulamentacao-da-inteligencia-artificial/. Acesso em: 5 fev. 2025.

HENDERSON, P. et al. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. **Journal of Machine Learning Research**, v. 21, p. 1-50, 2020.

ISO. ISO/IEC 42001: Artificial Intelligence Management System – Requirements. **International Organization for Standardization,** 2023.

JACKSON, N. Google's and Microsoft's Race Towards Carbon-Neutral AI. **Sustainable Computing Journal**, v. 12, p. 45-60, 2021.

JAISWAL, S. Regulamentação da IA: Entendendo as políticas globais e seu impacto nos negócios. Disponível em: https://www.datacamp.com/pt/blog/ai-regulation?utm_source=chatgpt.com. Acesso em: 5 fev. 2025.

JIANG, F. et al. Artificial intelligence in healthcare: past, present and future. **Stroke and Vascular Neurology**, v. 2, n. 4, p. 230-243, 2017.

JOBIN, A.; IENCA, M.; VAYENA, E. The global landscape of AI ethics guidelines. **Nature Machine Intelligence**, v. 1, n. 9, p. 389-399, 2019.

JO, H. et al. ***On the Robustness of Neural Networks to Adversarial Examples***. 2019. Disponível em: https://arxiv.org/abs/1903.08058. Acesso em: 06 fev. 2025.

JONES, B.; KINNEY, D. Government Support for AI Adoption in SMEs. **Journal of Economic Policy and Technology**, v. 19, n. 3, p. 78-94, 2023.

KATZENBACH, C.; ULBRICHT, L. Algorithmic governance: The regulation of algorithms in a digital society. **Big Data & Society**, v. 6, n. 1, p. 1-11, 2019.

KIRÁLY, G.; KERÉNYI, A. Artificial Intelligence in Banking: Opportunities and Challenges. **Journal of Financial Services Research**, v. 59, p. 147-167, 2021.

KOTSIANTIS, S. B. et al. Supervised machine learning: A review of classification techniques. **Informatica**, v. 31, p. 249-268, 2007.

KOUROU, T. et al. Machine learning applications in cancer prognosis and prediction. **Computational and Structural Biotechnology Journal**, v. 13, p. 8-17, 2015.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ***Imagenet classification with deep convolutional neural networks*****.** 2012. Disponível em: https://www.cs.toronto.edu/~fritz/absps/imagenet.pdf. Acesso em: 06 fev. 2025.

KUMAR, A.; RAJAN, P.; SING, A. Artificial Intelligence in Finance: Risk and Regulatory Challenges. **Journal of Financial Regulation and Compliance**, n. 29(4), p. 315-332, 2021.

KUMAR, R.; KUMAR, V. Hybrid AI Models in Business Decision-Making. **Journal of Management Analytics**, v. 9, p. 125-142, 2022.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015.

LI, X. et al. Computing Power and AI Adoption in Small Firms. **Journal of Emerging Technologies**, v. 8, n. 4, p. 201-225, 2022.

LIANG, X.; LÜTHJE, B. The Chinese digital economy: Revolution or evolution?. **New Left Review**, v. 134, p. 75-99, 2022.

LIPTON, Z. C. The mythos of model interpretability. **arXiv preprint arXiv:1606.03490**, 2018.

LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical Image Analysis**, v. 42, p. 60-88, 2017.

LUNDERVOLD, A.; LUNDERVOLD, A. An Overview of Deep Learning in Medical Imaging. **Frontiers in Neuroscience**, v. 13, p. 1-22, 2019.

MCCARTHY, J. et al. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. **AI Magazine**, v. 27, n. 4, p. 12-14, 1956.

MCKINNEY, S. M. et al. International evaluation of an AI system for breast cancer screening. **Nature**, v. 577, p. 89-94, 2020.

MICKOLEIT, A. Artificial Intelligence, Ethics, and Regulation: Global Challenges and Perspectives. **International Journal of Ethics and Technology**, v. 5, p. 32-47, 2021.

MITCHELL, M. et al. Addressing algorithmic bias in artificial intelligence. **Communications of the ACM**, v. 63, n. 10, p. 14-21, 2020.

MITCHELL, M. et al. Model Cards for Model Reporting. **Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)**, p. 220-229, 2021.

MORLEY, J. et al. Ethics of AI in health care: a mapping review. **Social Science & Medicine**, v. 260, p. 113172, 2020.

MOROZOV, E. To save everything, click here: The folly of technological solutionism. **PublicAffairs**, 2013.

MURPHY, K. P. Machine Learning: A Probabilistic Perspective. **MIT Press**, 2012.

NARAYANAN, A.; SHMATIKOV, V. Robust De-anonymization of Large Sparse Datasets. **IEEE Symposium on Security and Privacy**, 2008.

NGUYEN, T. H.; SHIRAI, K.; VELCIN, J. Sentiment analysis on social media for stock movement prediction. ***Expert Systems with Applications***, v. 42, n. 24, p. 9603-9611, 2015.

NOBLE, S. U. Algorithms of oppression: How search engines reinforce racism. **NYU Press**, 2018.

OBERMEYER, Z. et al. Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, v. 366, n. 6464, p. 447-453, 2019.

OECD. Recommendation of the Council on Artificial Intelligence. **OECD Publishing**, 2019. Disponível em: https://www.oecd.org/going-digital/ai/principles. Acesso em: 5 fev. 2025.

PAPERNOT, N. et al. ***Towards evaluating the robustness of neural networks*****.** 2017. Disponível em: https://arxiv.org/abs/1608.04644. Acesso em: 06 fev. 2025.

PATTERSON, D. et al. Carbon Emissions and Large Neural Network

Training. **Proceedings of the Conference on Sustainable Computing**, 2021.

PECK, J. et al. ***The role of machine learning in the analysis of large-scale data***. 2019. Disponível em: https://arxiv.org/abs/1906.07785. Acesso em: 06 fev. 2025.

PEGORARO, A. O que é corrupção e quais são os tipos? **Kronoos.com**, 12 out. 2023.

PRICE, W. N. Big data and black-box medical algorithms. **Science Translational Medicine**, v. 9, n. 448, p. eaao5333, 2017.

RAJI, I. D. et al. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. **FAT**, 2020.

RAJKOMAR, A. et al. Scalable and accurate deep learning with electronic health records. **npj Digital Medicine**, v. 1, p. 18, 2018.

RAMSAY, J. et al. ***The case for automated machine learning***. 2019. Disponível em: https://arxiv.org/abs/1901.06710. Acesso em: 06 fev. 2025.

RAO, A.; VERWEIJ, G. Talent Scarcity in the AI Industry: A Barrier for Small Enterprises. **Global Tech Trends**, v. 9, n. 2, p. 159-178, 2021.

RAU, M. et al. ***Blockchain-Based Decentralized Applications: A Survey***. 2020. Disponível em: https://arxiv.org/abs/2003.01891. Acesso em: 06 fev. 2025.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. “Why Should I Trust You?”

Explaining the Predictions of Any Classifier. **Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)**, p. 1135-1144, 2016.

ROY, M. C. O'N. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. **New York: Crown Publishers**, 2016. 272p. Hardcover, $26 (ISBN 978-0553418811). **College & Research Libraries**, v. 78, n. 3, p. 403, 19 abr. 2017.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: A Modern Approach. 4. ed. **Upper Saddle River: Pearson**, 2021.

SAMEK, W. et al. ***Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models***. 2017. Disponível em: https://arxiv.org/abs/1708.08296. Acesso em: 06 fev. 2025.

SANTANA, J. et al. ***IBM Watson for Oncology: A ferramenta que combina Inteligência Artificial e experiência clínica para tratamentos personalizados*****.** 2021.

SHEN, J. et al. An overview of artificial intelligence applications in healthcare. **npj Digital Medicine**, v. 4, n. 1, p. 1-14, 2021.

SHIN, E.; PARK, J. ***Decentralized Technologies and their Impact on Data Management in Cloud Computing*****.** 2019. Disponível em: https://www.researchgate.net/publication/332434589. Acesso em: 06 fev. 2025.

SHORTLIFFE, E. H. *Computer-Based Medical Consultations:* ***MYCIN.*** New York: Elsevier, 1976.

SHORTLIFFE, E. H.; SEPÚLVEDA, M. J. ***Clinical Decision Support Systems: Challenges and Opportunities*.** 2018. Disponível em: https://pubmed.ncbi.nlm.nih.gov/30240910/. Acesso em: 06 fev. 2025.

SICHMAN, J. S. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**, v. 35, n. 101, p. 37–50, abr. 2021.

SRNICEK, N. Platform capitalism. **Polity Press**, 2017.

STEINHUBL, S. R. et al. The emerging field of mobile health: research, technology, and policy. **Health Affairs**, v. 36, n. 2, p. 269-276, 2019.

STRUBELL, E.; GANESH, A.; MCCALLUM, A. Energy and Policy Considerations for Deep Learning in NLP. **Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)**, p. 3645-3650, 2019.

STUCKE, R. A.; GRUNES, A. D. ***Big Data and Competition Policy*.** 2016. Oxford: Oxford University Press.

SUN, J. et al. ***Deep learning for computer vision: A brief review***. 2017. Disponível em: https://arxiv.org/abs/1705.01682. Acesso em: 06 fev. 2025.

SUSSKIND, R.; SUSSKIND, D. The future of the professions: How technology will transform the work of human experts. **Oxford University Press**, 2015.

SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction.* 2. ed. Cambridge: **MIT Press**, 2018.

SZEGEDY, C. et al. ***Going deeper with convolutions***. 2014. Disponível em: https://arxiv.org/abs/1409.4842. Acesso em: 06 fev. 2025.

TOPOL, E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. **Basic Books**, 2019.

UNESCO. Recommendation on the Ethics of Artificial Intelligence. **Paris**, 2021.

VAN DIJCK, J.; POELL, T.; DE WAAL, M. The platform society: Public values in a connective world. **Oxford University Press**, 2018.

VASWANI, A. et al. ***Attention is all you need***. 2017. Disponível em: https://arxiv.org/abs/1706.03762. Acesso em: 06 fev. 2025.

VEALE, M.; BORGEFURST, A. Regulating AI: The Compliance Challenge for SMEs. **Law and Technology Review**, v. 16, n. 2, p. 122-144, 2021.

VINUESA, R. et al. The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. **Nature Communications**, v. 11, p. 1-10, 2020.

VOIGT, P.; VON DEM BUSSCHE, A. The EU general data protection regulation (GDPR). **Springer International Publishing**, 2017.

VOORVAAT, T. GDPR and AI: Navigating the Challenges of Data Protection and Machine Learning. **Computer Law & Security Review**, v. 36, p. 105-118, 2020.

VON ESCHENBACH, W. Trust in AI Decision-Making. **AI & Ethics**, v. 1, p. 23-39, 2021.

WEXLER, J. et al. The What-If Tool: Interactive Probing of Machine Learning Models. **IEEE Transactions on Visualization and Computer Graphics**, v. 26, n. 1, p. 56-65, 2019.

WHITE HOUSE. Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, 2022. Disponível em: https://www.whitehouse.gov/ostp/ai-bill-of-rights. Acesso em: 4 fev. 2025.

YANG, H. et al. ***Decentralized Finance (DeFi): A Comprehensive Survey***. 2021. Disponível em: https://arxiv.org/abs/2105.04735. Acesso em: 06 fev. 2025.

ZELLERS, R. et al. ***Defending against neural fake news***. 2019. Disponível em: https://arxiv.org/abs/1905.12616. Acesso em: 06 fev. 2025.

ZEMEL, R. et al. Learning Fair Representations. **Proceedings of the International Conference on Machine Learning (ICML)**, p. 325-333, 2013.

ZENG, J. China's digital authoritarianism: AI governance and policy implications. **International Affairs**, v. 97, n. 4, p. 1215-1235, 2021.

ZUBOFF, S. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. **PublicAffairs**, 2019.