



UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
Programa de Pós-Graduação
Educação em Direitos Humanos

VIÉS ALGORÍTMICO EM INSTITUIÇÕES FINANCEIRAS:
Desafios e Soluções sob a Perspectiva dos Direitos Humanos

Hiana Ferreira Santos

Diamantina
2024

Hiana Ferreira Santos

VIÉS ALGORÍTMICO EM INSTITUIÇÕES FINANCEIRAS
Desafios e Soluções sob a Perspectiva dos Direitos Humanos

Dissertação apresentada ao programa de Pós-Graduação em Educação em Direitos Humanos, como requisito para obtenção do título de Especialista.

Orientador: Prof. Me. Ricardo de Oliveira Brasil Costa

Diamantina
2024

Catálogo na fonte - Sisbi/UFVJM

S237v Ferreira Santos, Hiana

2024 Viés Algorítmico em Instituições Financeiras [manuscrito]
: Desafios e Soluções sob a Perspectiva dos Direitos Humanos
/ Hiana Ferreira Santos. -- Diamantina, 2024.
74 p. : il.

Orientador: Prof. Ricardo de Oliveira Brasil Costa.

Monografia (Especialização em Educação em Direitos Humanos) -- Universidade Federal dos Vales do Jequitinhonha e Mucuri, Curso de Especialização em Educação em Direitos Humanos, Diamantina, 2024.

1. Viés algorítmico. 2. Inteligência artificial. 3. Setor financeiro. 4. Direitos humanos. 5. Mitigação. I. Costa, Ricardo de Oliveira Brasil. II. Universidade Federal dos Vales do Jequitinhonha e Mucuri. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFVJM com os dados fornecidos pelo(a) autor(a).

Este produto é resultado do trabalho conjunto entre o bibliotecário Rodrigo Martins Cruz/CRB6-2886

e a equipe do setor Portal/Diretoria de Comunicação Social da UFVJM

Hiana Ferreira Santos

VIÉS ALGORÍTMICO EM INSTITUIÇÕES FINANCEIRAS

Desafios e Soluções sob a Perspectiva dos Direitos Humanos

Trabalho de Conclusão de Curso
apresentado ao Programa de Pós-
Graduação em Educação em Direitos
Humanos da Universidade Federal dos
Vales do Jequitinhonha e Mucuri, como
requisito parcial para a obtenção de
título de Especialista.

Orientador: Prof. Me. Ricardo Oliveira
Brasil Costa

Data de aprovação: 17/10/2024

Documento assinado digitalmente
 **RICARDO DE OLIVEIRA BRASIL COSTA**
Data: 22/10/2024 17:12:10-0300
Verifique em <https://validar.iti.gov.br>

Prof. Me Ricardo de Oliveira Brasil Costa Programa de Pós graduação Saúde
Sociedade e Ambiente - UFVJM

Documento assinado digitalmente
 **JHONNY MICHAEL COSTA**
Data: 21/10/2024 20:47:23-0300
Verifique em <https://validar.iti.gov.br>

Prof. Me. Jhonny Michael Costa
Diretoria de Educação Aberta e a Distância - EDH - UFVJM

Documento assinado digitalmente
 **RIANN MARTINELLI BATISTA**
Data: 22/10/2024 07:39:18-0300
Verifique em <https://validar.iti.gov.br>

Prof. Me. Riann Martinelli Batista
Diretoria de Educação Aberta e a Distância - EDH - UFVJM

Diamantina- MG

Aos meus queridos filhos *Danilo e Alice*, cuja existência ilumina meu caminho e inspira minha busca por um mundo mais *justo e igualitário*. Que este trabalho seja uma pequena contribuição para um futuro em que vocês possam viver em uma sociedade *mais humana, ética e responsável*. Dedico a vocês, com todo o amor e esperança.

AGRADECIMENTOS

Este trabalho não seria possível sem o apoio e a contribuição de muitas pessoas, às quais gostaria de expressar minha mais sincera gratidão.

Ao meu marido, cuja presença constante, amor e compreensão foram pilares inabaláveis durante todo esse processo. Sua parceria e apoio nos momentos mais difíceis me deram a força necessária para seguir adiante, e sua confiança em mim foi o combustível que me impulsionou a superar cada desafio.

Aos meus filhos, por serem a luz que guia meus passos e a razão pela qual continuo a acreditar em um futuro melhor. Seu carinho, paciência e energia renovam minhas forças quando o cansaço parecia dominar.

Ao meu orientador, professor Ricardo Brasil, por sua paciência, dedicação e pela confiança depositada em mim ao longo desta jornada. Sua orientação criteriosa e valiosas sugestões foram fundamentais para o desenvolvimento deste projeto, e sou profundamente grata por suas contribuições e por todo o aprendizado que me proporcionou.

Aos meus colegas do BRB, por compartilharem comigo a rotina diária de desafios e aprendizagens, sempre dispostos a ajudar e a motivar. O ambiente de camaradagem e incentivo que encontrei com vocês foi essencial para que eu pudesse conciliar minhas responsabilidades profissionais com a elaboração deste trabalho.

Aos meus amigos online do curso de Educação em Direitos Humanos, pela troca de conhecimentos, pelo suporte mútuo e pelas conversas enriquecedoras que me ajudaram a amadurecer diversas ideias e a olhar o mundo sob diferentes perspectivas. Agradeço a cada um de vocês pela amizade e pelo estímulo constante.

A todos, meu mais sincero agradecimento. Este trabalho é, em parte, fruto da colaboração, do carinho e do apoio de cada um de vocês.

"Negar às pessoas seus direitos humanos
é desafiar a própria humanidade."
— Nelson Mandela, *The Struggle Is My Life*
(1986)

RESUMO

Este trabalho tem como objetivo investigar o impacto do viés algorítmico no setor financeiro e suas implicações para os direitos humanos. O estudo aborda a interação entre Inteligência Artificial (IA) e Aprendizado de Máquina, identificando os principais tipos de viés, como o viés de dados, de modelo e de avaliação. A metodologia utilizada inclui uma revisão teórica sobre o viés algorítmico e sua relação com as tecnologias aplicadas ao setor financeiro, além de uma análise das iniciativas globais de mitigação do viés. Os resultados destacam como o viés algorítmico pode agravar a discriminação e a exclusão financeira, ampliando as desigualdades já existentes. O trabalho também propõe métodos de identificação e mitigação do viés, como o uso de algoritmos de equidade e a implementação de governança e *compliance* no desenvolvimento de sistemas de IA. Conclui-se que, embora existam desafios significativos, como a complexidade dos sistemas algorítmicos e a falta de regulação específica, a mitigação do viés algorítmico é fundamental para garantir o respeito aos direitos humanos no setor financeiro. O estudo recomenda a criação de novas políticas públicas e diretrizes para empresas e desenvolvedores, visando a construção de sistemas de IA mais justos e transparentes.

Palavras-chave: Viés algorítmico; Inteligência Artificial; Setor financeiro; Direitos humanos; Mitigação.

ABSTRACT

This study aims to investigate the impact of algorithmic bias in the financial sector and its implications for human rights. The research explores the interaction between Artificial Intelligence (AI) and Machine Learning, identifying the main types of bias, such as data bias, model bias, and evaluation bias. The methodology includes a theoretical review of algorithmic bias and its relationship with technologies applied to the financial sector, along with an analysis of global initiatives to mitigate bias. The results highlight how algorithmic bias can exacerbate discrimination and financial exclusion, thus amplifying existing inequalities. The study also proposes methods for identifying and mitigating bias, such as the use of fairness algorithms and the implementation of governance and compliance in AI system development. It concludes that, despite significant challenges, such as the complexity of algorithmic systems and the lack of specific regulations, mitigating algorithmic bias is essential to ensuring respect for human rights in the financial sector. The study recommends the creation of new public policies and guidelines for companies and developers to build fairer and more transparent AI systems.

Keywords: Algorithmic bias; Artificial Intelligence; Financial sector; Human rights; Mitigation.

LISTA DE ILUSTRAÇÕES

<u>Figura 1 - Desigualdade vs igualdade vs equidade vs justiça: uma explicação visual (GARCIA et. al, 2023)</u>	30
<u>Figura 2 - Viés no ciclo de vida do software (GARCIA et. al, 2023)</u>	32
<u>Figura 3 - Gastos além do orçamento (Banco Central do Brasil, 2017)</u>	44

LISTA DE ABREVIATURAS E SIGLAS

BRB	Banco de Brasília
DUDH	Declaração Universal dos Direitos Humanos
EDH	Educação em Direitos Humanos
ESG	<i>Environmental, Social, and Governance</i> (Ambiental, Social e Governança)
IA	Inteligência Artificial
ML	<i>Machine Learning</i> (Aprendizado de Máquina)
ODS	Objetivos de Desenvolvimento Sustentável
PNEDH	Programa Nacional de Educação em Direitos Humanos
TCC	Trabalho de Conclusão de Curso
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

LISTA DE SÍMBOLOS

© Copyright

Sumário

INTRODUÇÃO	17
1.1. Contextualização do Tema	18
1.2. Justificativa	18
1.3. Objetivos	22
1.3.1. Objetivo Geral	22
1.3.2. Objetivos Específicos	22
1.4. Metodologia	22
FUNDAMENTAÇÃO TEÓRICA	24
2.1. Inteligência Artificial e Aprendizado de Máquina	25
2.1.1. Conceitos Básicos	26
2.1.2. Tipos de Algoritmos	26
2.2. Viés Algorítmico	28
2.2.1. Definição e Tipos de Viés	30
2.2.2. Causas do Viés Algorítmico	31
2.3. Direitos Humanos e Tecnologia	33
RELAÇÃO ENTRE VIÉS ALGORÍTMICO E DIREITOS HUMANOS NO SETOR...	35
FINANCEIRO	35
3.1. Contexto do Setor Financeiro e Viés Algorítmico	36
3.1.1. Transformação Digital no Setor Financeiro	38
3.2. Impactos do Viés Algorítmico nos Direitos Humanos	39
3.2.1. Discriminação e Exclusão Financeira	41
3.2.2. Sustentabilidade Financeira e Responsabilidade Ética	45
MITIGAÇÃO DO VIÉS ALGORÍTMICO	46
4.1. Técnicas de Mitigação	47
4.2. Governança e <i>Compliance</i>	49
4.2.1. Política de Responsabilidade Social, Ambiental e Climática (PRSAC)	51
REGULAÇÃO E POLÍTICAS PÚBLICAS	53
5.1. Regulações Existentes	54
5.2. Propostas de Novas Políticas	55
5.2.1. Diretrizes para Empresas e Desenvolvedores	56

DISCUSSÃO.....	59
6.1. Futuras Tendências e Pesquisas.....	60
CONCLUSÃO	62
REFERÊNCIAS	64
GLOSSÁRIO	69

INTRODUÇÃO

1.1. Contextualização do Tema

No contexto atual, o avanço das tecnologias digitais tem transformado diversos setores, incluindo o financeiro. Contudo, o uso crescente de algoritmos e inteligência artificial tem levantado preocupações sobre possíveis vieses algorítmicos que podem impactar negativamente os direitos humanos.

De acordo com a UNESCO (2022), as tecnologias de IA têm o potencial de acentuar as divisões e desigualdades existentes tanto dentro dos países quanto entre eles. Portanto, é essencial promover justiça, confiança e equidade para garantir que nenhum país ou indivíduo seja excluído, seja no acesso justo às tecnologias de IA e seus benefícios, ou na proteção contra suas implicações negativas. Além disso, é importante reconhecer as diferentes circunstâncias de cada país e respeitar o desejo de algumas pessoas de não participar de todos os avanços tecnológicos.

A proliferação de tecnologias de inteligência artificial (IA) no setor financeiro trouxe consigo vantagens indiscutíveis, como a eficiência e a precisão nas operações financeiras. No entanto, conforme sublinhado pela UNESCO, a exploração dessas tecnologias não é isenta de riscos, sendo o viés algorítmico um dos mais preocupantes. Em vez de atuar como ferramentas neutras, algoritmos podem perpetuar preconceitos existentes e, em alguns casos, amplificar disparidades econômicas e sociais, causando sérios impactos sobre grupos marginalizados.

Quando esses sistemas são aplicados na concessão de crédito e avaliação de risco, podem resultar em padrões discriminatórios, estabelecendo uma barreira adicional para certos grupos obterem acesso a recursos financeiros. Isso é particularmente problemático em economias emergentes, onde as desigualdades são mais acentuadas e a proteção regulatória pode ser menos robusta.

Garcia et al. (2023) destacam que a aprovação de empréstimos depende de várias características dos tomadores, refletindo sua capacidade e disposição de pagar a dívida. Um dos fatores essenciais é o histórico de crédito do solicitante, que nem sempre está disponível, especialmente para imigrantes, estudantes, jovens profissionais e pessoas de baixa renda. Essa falta de informação tem sido um desafio para os bancos. No entanto, as *fintechs* estão abordando essa questão ao incluir outros tipos de dados, como o comportamento nas mídias sociais e os pagamentos de telecomunicações dos solicitantes.

De acordo com Garcia et al. (2023), a análise automatizada de solicitações de crédito reduz a subjetividade no processo de tomada de decisão. No entanto, como os algoritmos de *machine learning* são treinados com base em decisões históricas registradas nos conjuntos de dados das instituições financeiras, o processo tende a reforçar os preconceitos e vieses já existentes contra grupos definidos por raça, sexo, orientação sexual e outros atributos. Por essa razão, o interesse em identificar, prevenir e mitigar a discriminação algorítmica tem crescido significativamente em áreas como Ciência da Computação, Economia, Direito e Ciências Sociais.

Nesse contexto, investigar como esses vieses se manifestam e seus efeitos é crucial para a integridade do sistema financeiro. A concessão de crédito, como citado acima, depende cada vez mais de algoritmos que analisam vastas quantidades de dados para decidir sobre a elegibilidade dos solicitantes. Se estes algoritmos apresentarem vieses, indivíduos de comunidades marginalizadas podem ser injustamente reprovados, agravando sua exclusão financeira. Da mesma forma, na avaliação de risco, vieses podem alterar negativamente a forma como diferentes grupos são percebidos em termos de solvência, induzindo discriminações sutis ou explícitas.

Lima et al. (2009) observam que, assim como outros mecanismos matemáticos, como a análise discriminante e a regressão logística, os modelos de redes neurais também podem cometer erros de classificação na análise de crédito. No entanto, apesar desses possíveis erros, os mecanismos automatizados proporcionam à empresa subsídios e agilidade na tomada de decisões.

A exploração das consequências intencionais e não intencionais dos algoritmos é necessária e oportuna, principalmente porque as políticas públicas atuais podem não ser suficientes para identificar, mitigar e remediar os impactos ao consumidor (Lee, Resnick & Barton, 2019).

A busca por uma IA financeira mais justa e inclusiva não é uma tarefa simples, mas é imprescindível. O avanço tecnológico deve ser acompanhado de uma responsabilidade social, garantindo que a automação e a inteligência nos serviços financeiros contribuam para a redução das desigualdades, e não para a sua ampliação. Ao entender e mitigar os vieses algorítmicos, podemos promover um sistema financeiro que seja verdadeiramente acessível e equitativo para todos.

Para Cozman & Kaufman (2022), a constatação do viés em um modelo, em geral, ocorre tardiamente, dificultando identificar retroativamente sua origem e,

consequentemente, as formas de eliminá-lo. Dada a crescente visibilidade dos efeitos danosos do viés contido nas decisões automatizadas por IA, particularmente as aplicações em campos sensíveis como saúde e educação, especialistas acadêmicos e não acadêmicos, das ciências exatas e das ciências sociais.

A regulamentação também desempenha um papel crucial na mitigação do viés algorítmico. Leis e diretrizes específicas devem ser estabelecidas para garantir que as decisões automatizadas sejam justas e imparciais. Essas regulamentações devem prever mecanismos de fiscalização e penalidades para assegurar a conformidade. Instituições financeiras devem ser obrigadas a realizar auditorias regulares e independentes de seus sistemas automatizados para garantir a equidade em suas operações.

Neste trabalho, abordaremos, com enfoque especial, a relação entre o viés algorítmico e as diretrizes estabelecidas pela Declaração Universal dos Direitos Humanos (DUDH) e as políticas de Responsabilidade Socioambiental e de Governança Corporativa (PRSAC). A relação entre o viés algorítmico e as diretrizes pactuadas nas políticas de Responsabilidade Socioambiental e de Governança Corporativa é de crucial importância no cenário atual de avanços tecnológicos e digitais.

Em suma, o viés algorítmico pode representar uma ameaça real aos direitos humanos no setor financeiro, mas estratégias bem delineadas podem mitigar esses efeitos. Ao adotar uma abordagem integrada que combina soluções técnicas, regulamentação eficaz e princípios éticos, é possível promover a justiça social e assegurar que os avanços tecnológicos beneficiem a todos, sem discriminação.

1.2. Justificativa

A escolha do tema deste Trabalho de Conclusão de Curso (TCC) está embasada nos princípios fundamentais do Programa Nacional de Educação em Direitos Humanos (PNEDH, BRASIL, 2007), os quais promovem a igualdade, transparência, responsabilidade ética, inclusão social e a proteção dos direitos econômicos e sociais.

Primeiramente, a promoção da igualdade e da não discriminação é um pilar destacado pelo PNEDH. Estudar o viés algorítmico no setor financeiro pode revelar como sistemas automatizados perpetuam ou amplificam desigualdades sociais, econômicas e raciais. Esta análise é essencial para desenvolver soluções que

assegurem um tratamento justo e igualitário para todas as pessoas, independentemente de sua origem ou condição.

No que tange ao direito à informação e à transparência, outro pilar do PNEDH, a análise dos algoritmos financeiros sob a ótica dos direitos humanos visa aumentar a clareza desses sistemas. Isso possibilita que os indivíduos compreendam como suas informações são usadas e como influenciam decisões financeiras que impactam diretamente suas vidas. A transparência no tratamento de dados promovida por este TCC é um passo importante para a construção de um sistema financeiro mais justo.

A responsabilidade social e a ética no desenvolvimento e aplicação de novas tecnologias são também enfatizadas pelo PNEDH. Este estudo, ao examinar as implicações éticas do uso de algoritmos enviesados no setor financeiro, alinha-se com esses princípios. Ao promover um debate sobre a responsabilidade dos desenvolvedores e das instituições financeiras, o trabalho incentiva a adoção de práticas mais éticas e responsáveis.

A inclusão e participação social são fundamentais para o PNEDH. Avaliar como o viés algorítmico pode excluir certos grupos de oportunidades financeiras ou tratamento justo é vital para garantir que todos, sem exceção, tenham acesso equitativo a recursos financeiros e possam participar plenamente da vida econômica.

O impacto nos direitos econômicos e sociais de indivíduos é uma área de grande importância dentro do PNEDH. Este TCC destaca como o viés algorítmico em decisões financeiras pode afetar o acesso ao crédito, oportunidades de emprego e outros aspectos cruciais da vida econômica. Assim, comprometer estes direitos econômicos é uma questão que merece atenção e análise detalhada para buscar soluções viáveis.

A conscientização e educação em direitos humanos são objetivos centrais do PNEDH. Desenvolver um TCC sobre este tema contribui significativamente para ampliar o conhecimento sobre a influência das tecnologias nos direitos humanos. Esta abordagem promove um debate mais informado e crítico sobre as práticas atuais no setor financeiro, essencial para a evolução de políticas e práticas mais justas e éticas.

Finalmente, o desenvolvimento de políticas públicas que protejam os direitos humanos é incentivado pelo PNEDH. Identificar e propor soluções para mitigar o viés algorítmico pode fornecer entendimentos valiosos para a formulação de políticas públicas. Assim, este estudo contribui para a construção de um sistema financeiro

mais justo e equitativo, alinhando-se com o objetivo de proteger e promover os direitos humanos na esfera econômica e social.

1.3. Objetivos

1.3.1. Objetivo Geral

Investigar e propor estratégias para mitigar o viés algorítmico no setor financeiro, buscando que as práticas de modelagem e aplicação de algoritmos estejam alinhadas com os princípios de justiça, equidade, e direitos humanos, a fim de promover um sistema financeiro mais inclusivo e ético.

1.3.2. Objetivos Específicos

- Compreender os conceitos básicos de algoritmo, viés e suas implicações nos direitos humanos.
- Estabelecer a relação entre os direitos humanos e o viés algorítmico no setor financeiro, com foco no contexto nacional.
- Propor práticas para mitigar a influência dos dados processados por algoritmos nos direitos humanos.
- Avaliar as implicações legais e éticas relacionadas ao viés algorítmico no setor financeiro.

1.4. Metodologia

Na seção de metodologia deste trabalho, adotou-se uma abordagem qualitativa, baseada em revisão bibliográfica e análise documental. Inicialmente, foi realizada uma revisão da literatura sobre Inteligência Artificial (IA), Aprendizado de Máquina e viés algorítmico, com o intuito de compreender os conceitos fundamentais e identificar os principais tipos de viés que afetam sistemas de IA. Em seguida, foram analisadas as legislações e diretrizes existentes, tanto nacionais quanto internacionais, relacionadas à governança de IA e ao combate ao viés algorítmico, especialmente no setor financeiro.

Na análise realizada, foram inicialmente revisados 45 artigos científicos, dos quais 12 foram excluídos por não atenderem aos critérios estabelecidos, como a falta de relevância direta com o tema de viés algorítmico ou inconsistências metodológicas. A inclusão dos 34 artigos restantes se baseou na pertinência ao contexto de direitos humanos e discriminação algorítmica, bem como na robustez de suas metodologias. Além disso, a pesquisa contou com a análise de 8 legislações e 11 dissertações e teses, que auxiliaram na construção de um embasamento teórico e legal adequado para o tema investigado.

A metodologia também incluiu a avaliação de estudos de caso e iniciativas globais que visam mitigar o viés em sistemas algorítmicos, com foco em boas práticas e resultados obtidos por empresas e organizações. Além disso, foi feita uma análise crítica das políticas públicas e propostas regulatórias voltadas para o controle e mitigação do viés algorítmico, destacando os desafios e as oportunidades para o setor financeiro.

Por fim, o trabalho se baseou em uma análise descritiva dos impactos do viés algorítmico sobre os direitos humanos no contexto financeiro, considerando aspectos como discriminação, exclusão e desigualdade de oportunidades. O levantamento teórico e documental foi complementado com recomendações práticas para mitigação do viés, visando garantir maior equidade nos sistemas de IA utilizados por instituições financeiras.

FUNDAMENTAÇÃO TEÓRICA

2.1. Inteligência Artificial e Aprendizado de Máquina

A pergunta "As máquinas podem pensar?" é uma provocação filosófica que remonta ao famoso teste de Turing, proposto pelo cientista da computação Alan Turing em 1950 (Turing, 1950). O teste de Turing, também conhecido como jogo da imitação, foi desenvolvido para explorar os limites da inteligência artificial e se uma máquina pode exibir comportamento inteligente indistinguível do de um ser humano.

No teste original proposto por Turing (1950), participam três agentes: um interrogador e dois respondentes. O interrogador faz perguntas aos dois respondentes, sem saber qual é o humano e qual é a máquina. O objetivo do teste é verificar se o interrogador consegue distinguir entre respostas humanas e máquinas. Se o interrogador não conseguir diferenciar, a máquina passa no teste, sugerindo que ela pode "pensar".

Embora o teste de Turing tenha sido uma abordagem inovadora para a questão da inteligência artificial, ele levanta debates complexos sobre o que significa realmente "pensar". Por um lado, o sucesso no teste pode indicar que uma máquina é capaz de simular processos mentais humanos de maneira convincente. Por outro lado, críticos argumentam que simular um comportamento não é o mesmo que possuir consciência ou compreensão genuína.

A questão "As máquinas podem pensar?" se desdobra em questões secundárias relacionadas à consciência, entendimento, e até mesmo ética. A capacidade de uma máquina responder de maneira indistinguível de um humano não implica necessariamente que a máquina tenha consciência ou experiências subjetivas. Portanto, enquanto o teste de Turing é uma ferramenta valiosa para medir certos aspectos da inteligência artificial, ele não esgota a complexidade da questão sobre a possibilidade de máquinas pensarem.

A inteligência artificial (IA) e o aprendizado de máquina (ML) formam a espinha dorsal das inovações tecnológicas do futuro. A IA é a habilidade das máquinas de executar tarefas que requerem inteligência humana, como percepção visual, reconhecimento de fala e tomada de decisões complexas. Para Gonzales (2023), em contraste com as tecnologias tradicionais que automatizam processos de rotina, as tecnologias de IA vão ainda mais longe para imitar atividades humanas que exigem cognição, e sua aplicação e uso não se limitam ao processo de produção.

Gonzales (2023) ainda complementa com as contribuições da IA para o setor financeiro que incluem aplicações para detecção de fraudes e combate à lavagem de dinheiro. Além disso, desenvolvimentos de IA como “*robo-advice*” tornam “soluções de investimento personalizadas” possíveis no gerenciamento de metas financeiras e na otimização dos fundos dos clientes.

2.1.1. Conceitos Básicos

A Inteligência Artificial (IA) possui várias definições que se refere à capacidade das máquinas de realizarem tarefas que, quando realizadas por seres humanos, requerem inteligência, a União Internacional de Telecomunicações (UIT) define amplamente a IA como “sistemas adaptativos de autoaprendizagem”. Estas tarefas incluem a compreensão da linguagem natural, reconhecimento de padrões, resolução de problemas e tomada de decisões. Como apontado por Russell e Norvig (2016), “IA é o campo que estuda como construir sistemas computacionais que realizam tarefas que, em circunstâncias normais, exigiriam inteligência humana”.

O Aprendizado de Máquina é um subcampo da Inteligência Artificial que se concentra na construção de sistemas que podem aprender e melhorar a partir de dados sem serem explicitamente programados para isso. Em vez de seguir regras codificadas por humanos, os algoritmos de ML usam dados para identificar padrões e tomar decisões ou previsões com base nesses padrões.

Existem vários tipos de aprendizado de máquina, incluindo aprendizado supervisionado, não supervisionado e por reforço. Segundo Mitchell (1997), “aprendizado de máquina é o estudo de algoritmos e modelos estatísticos que os sistemas de computador utilizam para realizar uma tarefa específica sem usar instruções explícitas, mas sim através de padrões e inferências”.

2.1.2. Tipos de Algoritmos

Segundo Lee et al. (2019), no aprendizado de máquina, algoritmos dependem de vários conjuntos de dados, ou dados de treinamento, que especificam quais são as saídas corretas para algumas pessoas ou objetos. A partir desses dados de treinamento, ele aprende um modelo que pode ser aplicado a outras pessoas ou objetos e fazer previsões sobre quais devem ser as saídas corretas para eles.

Os algoritmos de aprendizado de máquina podem ser categorizados em três principais tipos: supervisionado, não supervisionado e por reforço. (VERMA, 2023)

De forma sucinta, como explicado por Verma (2023), os algoritmos supervisionados requerem dados rotulados para treinamento¹. Esses dados rotulados são essenciais porque contêm informações de entrada e a saída esperada, permitindo ao modelo aprender a fazer previsões ou classificações precisas. Por exemplo, em um problema de classificação de *e-mails* como *spam* ou não *spam*, o algoritmo supervisionado aprenderia a partir de exemplos pré-classificados e, conseqüentemente, aplicaria esse aprendizado a novos e-mails.

Os hiperparâmetros dos modelos baseados em aprendizado supervisionado são configurados com intuito de calibrar seu nível de assertividade e precisão. Essas características estão associadas ao viés e variância do modelo. O viés está relacionado à capacidade do modelo se ajustar aos dados aos quais foram apresentados durante o treinamento. Já a variância é a variabilidade das previsões do modelo. (Almeida et al., 2020)

A partir disto, compreende-se que como um modelo, os parâmetros que induzem o viés podem ser configurados pelo desenvolvedor. Quando um modelo se ajusta excessivamente aos dados de treinamento, ele pode captar não apenas o padrão subjacente, mas também o ruído aleatório presente nesses dados. Esse fenômeno é conhecido como *overfitting*.

De acordo com Dos Santos (2019), outro problema durante o processo de aprendizado está associado à memorização, onde o algoritmo não consegue obter as informações necessárias para realizar a tarefa com boa precisão. Quando isso ocorre, diz-se que o algoritmo apresentou *underfitting*. Isso geralmente acontece porque o modelo é muito simples para representar a complexidade dos dados, levando a previsões imprecisas e alta taxa de erro.

O aprendizado não supervisionado, ao contrário do aprendizado supervisionado, realiza inferências com base em um conjunto de dados que não foi previamente rotulado, classificado ou categorizado. Segundo Almeida et al. (2020),

¹ Em inteligência artificial (IA), dados rotulados para treinamento são conjuntos de dados que foram anotados com informações específicas que ajudam os modelos de aprendizado de máquina a aprender e fazer previsões. Esses rótulos fornecem contexto e significado aos dados brutos, permitindo que os algoritmos identifiquem padrões e tomem decisões baseadas nesses dados.

Por exemplo, em um conjunto de dados de imagens, cada imagem pode ser rotulada com a categoria do objeto que ela contém, como “gato” ou “cachorro”. Em um conjunto de dados de texto, as frases podem ser rotuladas com sentimentos como “positivo” ou “negativo”. Esses rótulos são essenciais para o treinamento supervisionado, onde o modelo aprende a partir de exemplos rotulados para prever rótulos de novos dados não rotulados

esse tipo de aprendizado é amplamente utilizado para descobrir padrões ocultos nos dados. Os autores também destacam que, ao empregar técnicas de aprendizado não supervisionado, é possível identificar estruturas e relacionamentos subjacentes nos dados que não são imediatamente evidentes, facilitando a exploração e análise de grandes volumes de informações.

Por fim, temos o aprendizado por reforço, onde o modelo interage com o ambiente e aprende através de tentativa e erro, recebendo recompensas ou penalidades com base em suas ações. Este tipo de aprendizado é particularmente relevante em campos como robótica e jogos, onde o sistema deve tomar decisões sequenciais para alcançar um objetivo otimizado a longo prazo. Para Almeida et al. (2020), o ambiente é o local em que o agente pode interagir tomando suas decisões. A priori o agente não possui nenhuma informação a respeito do ambiente, mas ele vai o conhecendo no decorrer de suas experiências para evoluir seus estados.

Essas três abordagens possibilitam uma ampla gama de ferramentas para resolver problemas complexos e diversificados, sublinhando a versatilidade e o potencial transformador do aprendizado de máquina no avanço tecnológico e científico.

2.2. Viés Algorítmico

Uma questão crucial em qualquer tipo de algoritmo é o viés. Segundo Barocas et al. (2014) o viés algorítmico ocorre quando mostra que conjuntos de dados são tendenciosos em relação uns aos outros em um sentido estatístico e é um bom ponto de partida para investigar se esses vieses incluem estereótipos culturais. Esse viés pode ser introduzido por vários fatores, como a seleção dos dados de treinamento ou as técnicas de agrupamento utilizadas.

Mitchell (1997) diz que os algoritmos supervisionados dependem fortemente de dados rotulados, e o viés presente nesses dados pode se refletir diretamente nas decisões tomadas pelo modelo. Por outro lado, algoritmos não supervisionados podem introduzir viés através das técnicas de agrupamento e análise escolhidas.

As tecnologias e, especificamente, os algoritmos, desempenham um papel crucial na sociedade contemporânea. A presença de viés em modelos de aprendizado de máquina é uma consequência direta das suposições feitas durante a construção e treinamento do modelo. Um viés adequado pode melhorar a generalização, mas um viés inadequado pode levar a erros sistemáticos e falhas na previsão. Para Barocas

et al. (2023), “nem tudo são más notícias: o aprendizado de máquina pode, de fato, ajudar a mitigar vieses de medição.”

Uma das principais formas pelas quais o viés algorítmico impacta os direitos humanos é através da discriminação. Por exemplo, algoritmos utilizados em processos de recrutamento de emprego podem, inadvertidamente, privilegiar candidatos de um gênero ou raça específica, se os dados de treinamento, rotulados, contiverem esses preconceitos. Isso não apenas perpetua desigualdades existentes, mas também nega oportunidades justas a indivíduos qualificados, violando seu direito ao trabalho e à igualdade.

Barocas et al. (2014) argumentam que a presença de disparidades não necessariamente indica uma intenção deliberada do designer do sistema em criar tais desigualdades. É crucial analisar além das intenções para compreender quando essas disparidades podem ser vistas como discriminação. Para isso, é importante questionar se as disparidades são justificadas e se causam prejuízos. Embora essas questões sejam complexas e raramente tenham respostas simples, a vasta literatura sobre discriminação nas áreas de filosofia e sociologia pode oferecer percepção para essa análise.

Segundo o que é entendido do texto de Barocas et al. (2014), entender e resolver o viés algorítmico exige uma análise detalhada dos dados e dos processos de treinamento empregados no desenvolvimento desses sistemas. Desenvolver algoritmos que promovam a justiça e a igualdade requer um esforço contínuo para garantir que os dados sejam representativos e que as disparidades injustificadas sejam corrigidas antes de afetarem os resultados. Dessa forma, podemos trabalhar para um futuro em que a tecnologia contribua para a equidade e o respeito aos direitos humanos.

Garcia et al. (2023) afirmam que a análise automatizada de solicitações de crédito reduz a subjetividade na tomada de decisões. No entanto, como o aprendizado de máquina se baseia em decisões passadas registradas nos dados das instituições financeiras, ele tende a perpetuar preconceitos e vieses existentes contra grupos específicos, como aqueles definidos por raça, sexo e orientação sexual. Por isso, o interesse em identificar, prevenir e mitigar a discriminação algorítmica tem crescido significativamente em diversas áreas, incluindo Ciência da Computação, Economia, Direito e Ciências Sociais.

A conscientização e a compreensão completa do viés algorítmico podem-se inferir que são essenciais para proteger os princípios dos direitos humanos. Implementar medidas de mitigação, como a revisão contínua dos algoritmos, a auditoria de dados para vieses e a promoção da diversidade e inclusão na equipe de desenvolvimento, podem ser passos essenciais. Apenas através de uma abordagem consciente e ética podemos buscar que a tecnologia alavanque direitos humanos em vez de comprometê-los.

2.2.1. Definição e Tipos de Viés

Igualdade implica fornecer às pessoas os mesmos recursos. Equidade significa fornecer às pessoas o número de recursos que cada uma precisa para atingir seu objetivo. Justiça se refere a fornecer às pessoas meios para que todas tenham a mesma oportunidade de atingir seus objetivos. (GARCIA et. al, 2023)



Figura 1 - Desigualdade vs igualdade vs equidade vs justiça: uma explicação visual (GARCIA et. al, 2023)

Vieses afetam diretamente a equidade e a justiça dos resultados, trazendo impactos negativos para determinados grupos ou indivíduos. É fundamental entender as diferentes formas como o viés pode surgir e como ele influencia os resultados dos sistemas de IA no dia a dia, que os vieses de dados, de modelo e de avaliação.

O viés de dados acontece quando os dados usados para treinar um modelo de IA não são representativos da população do mundo real. Isso pode ocorrer por diversas razões, como a coleta de dados de fontes limitadas ou a presença de amostras desbalanceadas. De acordo com Cozman e Kaufman (2022), o viés associado aos dados pode ter duas principais origens: os dados coletados podem não representar a composição proporcional do universo objeto em questão, ou podem refletir os preconceitos existentes na sociedade. O primeiro caso pode ocorrer, por exemplo, se uma base de dados de treinamento contiver mais observações de uma categoria que, de fato, é minoritária.

O viés de modelo é outro tipo de viés algorítmico que se refere a preconceitos incorporados durante o design e a arquitetura do próprio modelo de IA. Mesmo que os dados de treinamento sejam representativos, as escolhas feitas durante a construção do modelo podem introduzir vieses. Cozman e Kaufman (2022) destacam

que, entre as externalidades negativas, estão a opacidade dos modelos — ou seja, a falta de explicabilidade em como os algoritmos chegaram a determinadas saídas com base nos dados de entrada — e o viés contido nos resultados dos modelos. A constatação do viés, em geral, ocorre tardiamente, o que dificulta eliminá-lo e identificar retroativamente sua origem.

O viés de avaliação ocorre quando os critérios utilizados para avaliar o desempenho de um sistema de IA são tendenciosos, o que pode resultar em favorecimento de determinados grupos ou na exclusão de variáveis importantes. Isso pode ser observado, por exemplo, em sistemas de avaliação educacional que beneficiam alunos de determinados grupos socioeconômicos, exacerbando desigualdades existentes no sistema educativo. Esse tipo de viés é particularmente problemático, pois muitas vezes só é percebido após a ampla implementação do sistema. (Barocas et al., 2023)

A mitigação do viés algorítmico, como veremos adiante, é um desafio contínuo no campo da inteligência artificial. Exige uma abordagem integrada que considere não apenas a qualidade e a representatividade dos dados, mas também as escolhas arquiteturais e os métodos de avaliação. Ao reconhecer e abordar esses vieses, podemos desenvolver sistemas de IA mais justos e equitativos, que beneficiem a sociedade como um todo.

2.2.2. Causas do Viés Algorítmico

O viés pode se infiltrar durante todas as fases de um projeto, “...seja especificando o problema a ser resolvido de maneiras que afetam as classes de forma diferente, falhando em reconhecer ou abordar vieses estatísticos, reproduzindo preconceitos passados ou considerando um conjunto insuficientemente rico de fatores.” (Lee et al., 2019 apud Barocas et al., 2014)

“Os preconceitos humanos históricos são moldados por preconceitos generalizados e muitas vezes profundamente enraizados contra certos grupos, o que pode levar à sua reprodução e amplificação em modelos de computador.” (Lee et al., 2019) Se os dados forem coletados de forma desigual ou refletirem preconceitos históricos, o algoritmo tende a perpetuar esses vieses. Por exemplo, se um conjunto de dados usado para treinar um modelo de contratação contém principalmente informações sobre candidatos de um determinado grupo demográfico, o algoritmo pode aprender a preferir candidatos com características semelhantes.

Turner Lee (Lee et al., 2019) argumentou que muitas vezes é a falta de diversidade entre os programadores que projetam a amostra de treinamento que pode levar à sub-representação de um grupo específico ou de atributos físicos específicos. Muitas vezes, os engenheiros e desenvolvedores podem, inconscientemente, inserir seus próprios preconceitos no sistema. Isso pode ocorrer através de uma escolha inadequada das variáveis a serem consideradas, ou pelo ajuste de parâmetros que acabam favorecendo um determinado grupo.

A falta de diversidade nas equipes de desenvolvimento também contribui significativamente para esse problema. Equipes homogêneas são mais propensas a criar soluções que funcionam bem para indivíduos semelhantes a elas mesmas, negligenciando as necessidades e realidades de grupos menos representados. Portanto, promover a inclusão e a diversidade no ambiente de trabalho é essencial para reduzir o viés algorítmico.

Metodologias inadequadas de validação e teste também podem perpetuar vieses. Se os modelos não forem testados em dados suficientemente diversos, ou se as métricas de desempenho não incluírem medidas de justiça, o viés pode passar despercebido e ser incorporado nos resultados. Como podemos ver na imagem abaixo:

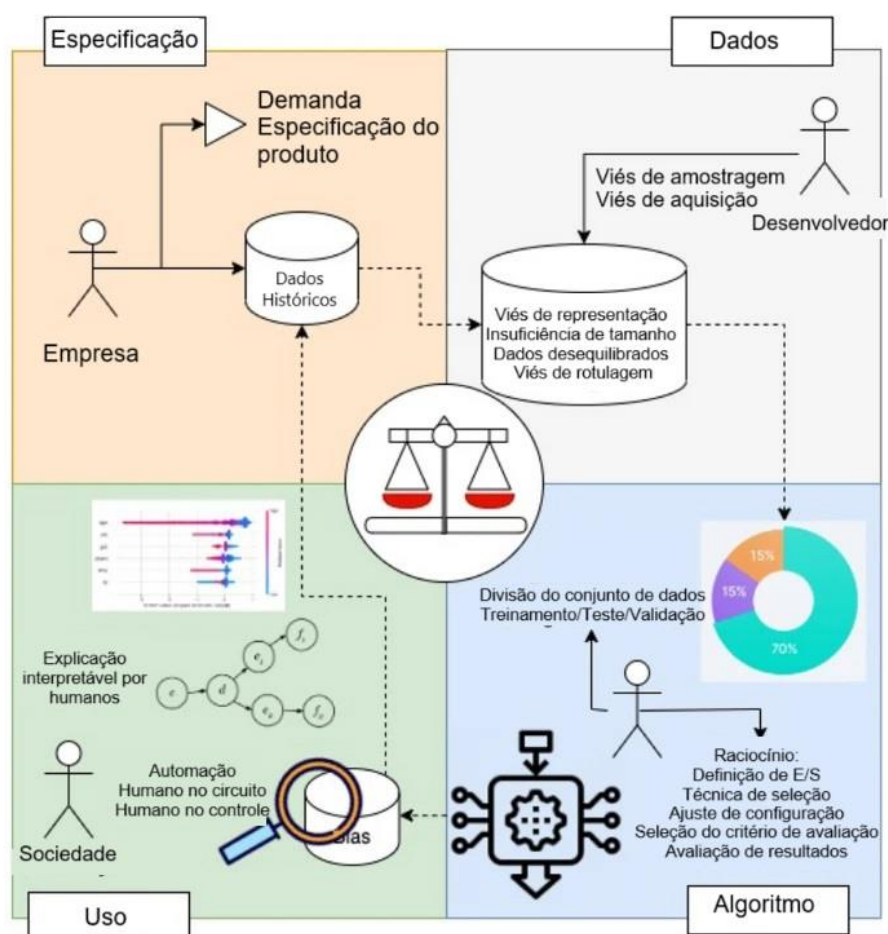


Figura 2 - Viés no ciclo de vida do software (GARCIA et. al, 2023)

Além disso, há questões relacionadas às limitações técnicas da própria inteligência artificial. Algoritmos de aprendizado de máquina muitas vezes funcionam como uma "caixa preta", onde não é claro como certas decisões são tomadas. Essa falta de transparência torna difícil identificar e corrigir vieses.

2.3. Direitos Humanos e Tecnologia

Enquanto por um lado, as tecnologias podem ser poderosas ferramentas para promover os direitos humanos, facilitando o acesso à informação, a comunicação e a organização social. Por outro lado, elas podem ser utilizadas para violar direitos fundamentais, como a privacidade, a liberdade de expressão e a não discriminação.

Goulart (2012) argumenta que, considerando que o artigo 27 da Declaração Universal dos Direitos do Homem assegura o “direito de participar livremente da vida cultural da comunidade” e de “participar do progresso científico”, o acesso à internet deve ser visto como um direito fundamental. Além disso, a liberdade de expressão é essencial para permitir a livre manifestação do pensamento e a comunicação na Internet. Portanto, países que restringem a manifestação de pensamento online estão desalinhados com o padrão mundial de liberdade de expressão nesse contexto.

Um dos principais desafios é o surgimento de novas formas de vigilância e controle, possibilitadas por tecnologias como o reconhecimento facial e a análise de dados em grande escala. A coleta indiscriminada de dados pessoais e a criação de perfis individuais podem ser utilizadas para fins de manipulação, discriminação e repressão. Segundo Ghisleni (p. 10, 2015) “o fluxo de informações pessoais foge do controle de seu titular, colocando-o em uma situação de vulnerabilidade frente às empresas ou ao próprio governo, principais praticantes da coleta massiva de dados pessoais.”

A inteligência artificial, por sua vez, apresenta tanto oportunidades quanto riscos. Embora a IA possa ser utilizada para resolver problemas complexos e melhorar a qualidade de vida, ela também pode perpetuar e amplificar vieses existentes, discriminando grupos minoritários.

Ghisleni (2015) destaca que, embora as novas capacidades tecnológicas tenham um potencial extraordinário para aplicações positivas, como melhorias no tráfego público, prevenção de crimes e combinação de ofertas de serviços, a coleta de dados apresenta riscos. Cada clique, login e download gera gigabytes de informações, muitas vezes fornecidas de forma consciente ou inconsciente, o que pode levar a diversas consequências sociais negativas devido ao uso inadequado desses dados.

De acordo com Floridi e Cowls (2019, apud De Jesus et al., 2024) desenvolvimento de sistemas de IA deve ser guiado por princípios éticos sólidos para garantir que essas tecnologias sejam utilizadas de forma equitativa e consciente, a ausência de regulamentações e diretrizes éticas bem definidas pode levar ao uso inadequado da IA, causando impactos negativos na privacidade, na segurança e equidade.

**RELAÇÃO ENTRE VIÉS ALGORÍTMICO E DIREITOS HUMANOS NO SETOR
FINANCEIRO**

3.1. Contexto do Setor Financeiro e Viés Algorítmico

Como visto anteriormente, o viés pode surgir de várias formas: na coleta de dados inicial, no design do algoritmo ou na interpretação dos resultados. No setor financeiro, isso se manifesta em áreas como concessão de crédito, definição de juros e avaliação de riscos. Por exemplo, se um algoritmo for treinado com dados históricos que refletem discriminação racial ou socioeconômica, ele pode perpetuar essas desigualdades, negando crédito a grupos já marginalizados.

É interessante observar a discriminação por duas óticas como feito por Garcia et. al. (2023), eles mostram que muitas empresas que vendem jogos exigem que os usuários registrados tenham mais de 18 anos [o que é uma discriminação, idadeismo]. Outro exemplo é negar crédito a pessoas que vivem em certas localidades [ele cita código postal, é um tipo de discriminação, xenofobia]. Enquanto o primeiro tipo de discriminação é aceitável, às vezes até exigido por lei, o último é ilegal. Quando uma discriminação é positiva ou não no tratamento de dados? Garcia et. al (2023) responde que, não importa quantas definições de justiça possamos chegar, elas permanecerão contestáveis em relação a algumas outras definições.

Essa situação é alarmante quando analisada sob a ótica dos direitos humanos. O princípio da igualdade e não discriminação é fundamental nesse campo. Se os algoritmos reforçam preconceitos implícitos, eles contribuem para a violação desses direitos, exacerbando desigualdades e promovendo injustiças. Além disso, a falta de transparência nos processos algorítmicos complica ainda mais a situação. Pessoas afetadas negativamente por decisões algorítmicas muitas vezes não têm como contestar ou sequer entender o processo por trás da decisão.

Gurgel (2007) afirma que o Princípio da Dignidade da Pessoa Humana, juntamente com os princípios da Igualdade e Não Discriminação e os Direitos Fundamentais, não podem ser hierarquizados, renunciados ou alienados. Esses princípios formam a base do respeito ao ser humano e fundamentam o conteúdo mínimo necessário para uma existência digna, que inclui o trabalho decente.

A responsabilidade pelas consequências do viés algorítmico não pode ser delegada exclusivamente às máquinas. Mas, como observado por Da Rocha et. al. (p. 12, 2020) “As *Big Techs* contam com a inteligibilidade do funcionamento algorítmico, não se responsabilizando pelos danos dos processos automatizados e invisíveis, como se vestissem uma capa de isenção por seus feitos.” Isso nos faz pensar que medidas como auditorias independentes, a inclusão de uma diversidade de dados no

treinamento e a implementação de protocolos de revisão ética são essenciais para mitigar o viés.

Da Rocha et al. (2020) argumentam que o lado escuro do *machine learning* que contribui para a incorporação de vieses racistas é que os algoritmos aprendem com dados representativos do mundo como ele é – profundamente impregnado de discriminação –, e não como ele deveria ser. Portanto, se os vieses não forem corrigidos nas raízes da programação, eles serão simplesmente automatizados. Para projetar algoritmos que promovam resultados mais justos, eles não devem refletir o legado de discriminações passadas e presentes, mas agir como um artifício não discriminatório.

Nos últimos anos, o setor financeiro tem passado por transformações significativas impulsionadas pelos avanços tecnológicos, mudanças regulatórias e dinâmicas econômicas globais. Como entendido no texto de Santos (2023), o advento dos canais digitais revolucionou a maneira como os serviços financeiros são prestados, introduzindo soluções mais ágeis e acessíveis para consumidores e empresas. Essas inovações têm desafiado os bancos tradicionais a repensarem seus modelos de negócios e a investirem em tecnologia para se manterem competitivos.

Para Carneiro (2002), “a regulação também desempenha um papel crucial na modelagem do setor”. Após a crise financeira de 2008, foram implementadas diversas medidas para aumentar a transparência e a segurança dos mercados financeiros. Regulamentos como Basileia III, MiFID II na Europa e Dodd-Frank nos Estados Unidos visam fortalecer a resiliência das instituições financeiras, exigindo capital mais robusto e melhores práticas de gestão de risco.

A globalização continua a pressionar o setor financeiro a se adaptar às realidades econômicas de diferentes regiões. Segundo Santos (2023) A sociedade digital demanda processos cada vez mais rápidos e inovadores. Com o objetivo de estimular e engajar os clientes por meio de um serviço diferenciado, houve adaptações. Com essa inovação, as leis e normas que regem o Sistema Financeiro Nacional foram atualizadas e modernizadas, buscando também garantir a segurança jurídica dos serviços financeiros. Além disso, preocupações crescentes com questões ambientais, sociais e de governança (ESG) estão influenciando investimentos e decisões de negócios, empurrando o setor em direção a práticas mais sustentáveis.

A digitalização não apenas transformou os serviços oferecidos, mas também criou desafios em termos de cibersegurança. Com o aumento das transações online

e o armazenamento de dados sensíveis, a proteção contra fraudes e ataques cibernéticos tornou-se uma prioridade para instituições financeiras de todos os tamanhos. Conforme Paloma Santos:

Santos (2023) destaca que a segurança continua sendo um aspecto crucial na escolha de serviços financeiros digitais. Para construir e manter a confiança dos clientes na proteção de seus dados pessoais e financeiros, os bancos digitais devem não apenas fortalecer suas medidas de segurança, mas também investir em educação contínua e comunicação eficaz. Isso é essencial para garantir que os clientes se sintam seguros e confiantes ao utilizar os serviços digitais oferecidos.

A pandemia de COVID-19 intensificou muitas dessas tendências, aumentando a dependência de tecnologia e a necessidade de resiliência operacional. Para Gonzaga (2023), diante da necessidade de rápida adaptação, muitas empresas passaram a implementar o trabalho remoto e soluções digitais para assegurar a continuidade das operações. A pandemia também destacou a relevância da inovação e da resiliência no mundo dos negócios.

O setor financeiro está em um estado constante de evolução, impulsionado por inovações tecnológicas, mudanças regulatórias, e pressões econômicas e sociais. Aqueles que conseguirem equilibrar esses fatores poderão não apenas sobreviver, mas prosperar em um ambiente cada vez mais competitivo e dinâmico.

3.1.1. Transformação Digital no Setor Financeiro

A transformação digital no setor financeiro tem sido um processo dinâmico e revolucionário, impulsionando grandes mudanças na forma como as instituições operam e interagem com seus clientes. A adoção de tecnologias emergentes, como inteligência artificial (IA), *blockchain* e *big data*, tem permitido às instituições financeiras oferecer serviços mais eficientes, personalizados e seguros.

Da Silva Caraffini et al. (2018) concluíram que, em uma pesquisa sobre a transformação digital em quatro instituições bancárias no Brasil, as principais iniciativas que suportam essa transformação estão focadas no fortalecimento dos canais digitais, como *mobile* e *internet banking*. Além disso, a promoção da inovação é incentivada através de espaços dedicados à implementação de novas ideias em parceria com entidades externas, e há um esforço contínuo para melhorar e otimizar os recursos internos, como a adoção de *cloud computing* e inteligência artificial.

Como citado por Da Silva Caraffini et. al. (2018), uma das principais áreas de impacto é a automação de processos. Com o uso de IA e *machine learning*, bancos e

outras instituições conseguem automatizar tarefas rotineiras, como análise de crédito e detecção de fraudes, reduzindo significativamente o tempo e os custos operacionais. Isso também libera os funcionários para se concentrarem em atividades de maior valor agregado, como o atendimento ao cliente e a elaboração de estratégias de negócios. As últimas tecnologias bancárias mencionadas na Pesquisa Febraban de Tecnologia Bancária (2024) incluem a Inteligência Artificial (IA) Generativa, usada para personalizar serviços e detectar fraudes, e a Computação em Nuvem (*Cloud*), adotada por todos os bancos, proporcionando escalabilidade e flexibilidade. Além disso, *Blockchain* e DREX, a moeda digital brasileira, e a Computação Quântica emergem como inovações que estão transformando a segurança e o poder de processamento no setor bancário.

Ainda, segundo a Pesquisa Febraban de Tecnologia Bancária (2024), 100% dos bancos já migraram suas operações para a nuvem, sendo a tecnologia cloud um dos principais pilares para agilidade e escalabilidade operacional. A inovação tecnológica no setor bancário está transformando as indústrias e cadeias produtivas, com investimentos crescentes em tecnologias como IA, computação quântica, e segurança cibernética, demonstrando o papel das instituições financeiras na liderança da transformação digital.

A transformação digital no setor financeiro pode estar redefinindo a maneira como as instituições operam e interagem com seus clientes, potencialmente oferecendo benefícios como automação, personalização e segurança. No entanto, enfrentar desafios como cibersegurança e regulação é considerado crucial para garantir uma implementação bem-sucedida dessa transformação. As instituições que abraçam a mudança e investem em tecnologia e talento podem estar mais bem posicionadas para prosperar no cenário financeiro em constante evolução.

3.2. Impactos do Viés Algorítmico nos Direitos Humanos

De acordo com Rossetti e Angeluci (2021), as habilidades de aprendizado adquirem aos algoritmos um certo nível de autonomia. O efeito dessa autonomia pode ser incerto, embora limitado. Por isso, as ações realizadas pelo aprendizado de máquina são difíceis de prever de antemão (como uma nova entrada será processada) ou de explicação depois (como uma decisão específica foi alcançada). Essa incerteza pode dificultar a identificação e a resolução de problemas éticos no desenvolvimento e na operação de algoritmos de aprendizagem, que possuem capacidade de tomar decisões.

Os autores Rossetti e Angeluci (2021) apresentam um mapa conceitual sobre a ética dos algoritmos, destacando sete questões éticas derivadas de preocupações

epistêmicas e normativas. Essas questões incluem a falibilidade do algoritmo, que pode levar a conclusões errôneas devido a correlações inadequadas; a opacidade, que impede a compreensão dos processos algorítmicos; o viés, que reflete preconceitos humanos; a discriminação, resultante da aplicação injusta de algoritmos; a autonomia, comprometida pela manipulação do comportamento do usuário; e a privacidade, ameaçada pela coleta de dados.

Rossetti & Angeluci (2021), enfatizam a responsabilidade ética associada a esses problemas, sugerindo a necessidade de um debate mais profundo sobre as implicações sociais e éticas dos algoritmos na tomada de decisões.

Ribeiro (2022) observa que, apesar das vantagens das decisões automatizadas, existe um risco inerente a essa abordagem. Como a sociedade é permeada por desigualdades, os modelos de *machine learning* podem “aprender” a reproduzir políticas discriminatórias presentes no ambiente socioeconômico. Sem medidas corretivas, isso pode gerar um círculo vicioso de retroalimentação, onde o sistema de inteligência artificial contribui para o agravamento dos problemas sociais.

Para adentrar nesse debate, é importante considerar que o viés algorítmico no setor financeiro pode representar um desafio significativo para os direitos humanos, afetando os princípios de explicabilidade, equidade e transparência. À medida que as instituições financeiras dependem cada vez mais de algoritmos para tomar decisões, como concessão de crédito e análise de risco, torna-se essencial compreender como esses sistemas podem perpetuar ou até exacerbar injustiças.

Os autores de Pádua & Lorenzetto (2024) definem a explicabilidade como a capacidade de descrever e justificar as decisões tomadas por sistemas de IA permitindo que sejam entendidas e confiáveis para os humanos. No contexto do setor financeiro, a falta de explicabilidade pode resultar em decisões que os clientes não conseguem contestar ou entender, violando seu direito à justiça e à igualdade. Por exemplo, um algoritmo que automaticamente nega um empréstimo sem fornecer uma justificativa clara impede que o cliente entenda ou conteste essa decisão. A falta de clareza sobre como os algoritmos chegam às suas conclusões dificulta a responsabilização das instituições financeiras e a implementação de mecanismos de recurso eficazes.

A equidade, ou *fairness*, é outro princípio fundamental que pode ser comprometido pelo viés algorítmico. Ribeiro (2022) explica que a área de pesquisa transdisciplinar “*machine learning and fairness*” ou “equidade algorítmica” sugere

técnicas e critérios para avaliar criticamente os algoritmos e dados, propondo soluções para minimizar os riscos de discriminação nas aplicações de dados. No setor financeiro, isso pode significar que certos grupos, como minorias raciais ou de gênero, sejam desfavorecidos nas decisões de crédito ou em outras áreas financeiras. Essa discriminação não apenas perpetua desigualdades sociais e econômicas, mas também representa uma violação direta dos direitos humanos, incluindo o direito à não discriminação.

A transparência é essencial para garantir que os sistemas algorítmicos sejam justos e responsáveis. De Pádua e Lorenzetto (2024) explicam que a transparência inclui oferecer acesso claro às informações e aos processos de tomada de decisão, detalhando como as decisões são tomadas e quais critérios e algoritmos (sequências de instruções lógicas) são utilizados. A falta de transparência dificulta a identificação e a correção de vieses algorítmicos, aumentando o risco de abusos dos direitos humanos. Uma maior transparência exigiria que as instituições financeiras não apenas divulgassem como seus algoritmos funcionam, mas também os parâmetros e dados utilizados, permitindo uma auditoria externa e independente das práticas algorítmicas.

A explicabilidade, equidade e transparência dos algoritmos podem ser fundamentais para mitigar os efeitos do risco algorítmico e garantir a proteção dos direitos humanos em conformidade com a Declaração Universal dos Direitos Humanos (DUDH). Quando algoritmos operam de forma opaca e discriminatória, há o risco de agravamento das desigualdades, violando direitos como igualdade perante a lei, acesso ao trabalho e condições justas, e até a segurança social. Para que a tecnologia sirva à justiça e à equidade, é importante promover o desenvolvimento de algoritmos explícitos e transparentes, com mecanismos que garantam a equidade e o respeito a todos os indivíduos.

3.2.1. Discriminação e Exclusão Financeira

A discriminação e exclusão financeira podem ser problemas interconectados que afetam milhões de pessoas ao redor do mundo. Enquanto a discriminação pode referir-se ao tratamento desfavorável de indivíduos com base em características como raça, gênero, orientação sexual ou status socioeconômico, a exclusão financeira pode dizer respeito à falta de acesso adequado a serviços financeiros essenciais, como contas bancárias, crédito, seguros e outros produtos financeiros.

Esses fenômenos têm um impacto significativo, perpetuando ciclos de pobreza e desigualdade. Rossetti e Angeluci (2021) destacam que os algoritmos

analisam dados, identificam padrões e aplicam esses resultados a pessoas específicas perfis detalhados aos padrões encontrados. O problema surge quando esses algoritmos discriminam com base em atributos sensíveis, como etnia, gênero ou orientação sexual, violando princípios fundamentais de equidade e transparência, e impactando diretamente os direitos humanos.

Garcia et. al. (2023) mencionam o estudo de Hassani (2021), que utilizou um conjunto de dados público do Kaggle para investigar como a aprovação de empréstimos bancários por meio de algoritmos pode perpetuar discriminação racial, étnica e de gênero. O conjunto de dados continha 400 instâncias, distribuídas por raça/etnia (99 “afro-americanos”, 102 “asiáticos”, 199 “caucasianos”) e sexo (207 mulheres e 193 homens), com 31% dos pedidos de empréstimo rejeitados. Usando as informações da pontuação de crédito, Hassani previu a raça e o sexo dos candidatos, e seus resultados indicaram que a discriminação racial e de gênero pode persistir nos empréstimos, mesmo após a remoção de atributos sensíveis dos dados (Garcia, ACB, Garcia, MGP & Rigobon, 2023, apud Hassani, 2021).

Crocco et. al. (2013), a definição de exclusão financeira está frequentemente associada a aspectos físicos e geográficos, o que, de certa forma, limita o conceito a práticas bancárias excludentes. Isso inclui, por exemplo, a localização de agências em áreas com níveis de renda mais elevados e a restrição de produtos financeiros a pessoas de baixa renda ou minorias sociais devido aos custos mais altos de monitoramento. Nesse sentido, os estudos sobre exclusão financeira têm expandido o foco, incluindo não apenas o acesso a serviços financeiros, mas também o seu uso e as consequências sociais.

Conforme Crocco et. al. (2013), a exclusão financeira envolve as dificuldades enfrentadas por situações de baixa renda para acesso a serviços financeiros essenciais, como abertura de contas, acesso a crédito e opções de pagamento sem o uso de dinheiro. Além disso, há a necessidade de criar portfólios de ativos com condições flexíveis, considerando também benefícios de segurança social e instabilidade no trabalho.

Mulheres e minorias raciais podem enfrentar obstáculos adicionais ao tentar obter empréstimos ou abrir contas bancárias, devido a preconceitos enraizados e práticas injustas. A ausência de histórico financeiro ou a desconfiança por parte dos bancos tradicionais também pode agravar a situação.

De acordo com o Banco Central do Brasil (2017), dados recentes mostram que ainda há muito a avançar em relação à participação econômica feminina no Brasil. O Fórum Econômico Mundial coloca o país na 90ª posição no ranking de desigualdade de gênero entre 142 países. Especificamente na dimensão de Participação Econômica e Oportunidade, o Brasil ocupa a 83ª posição, ficando atrás de outros países da América do Sul, como Colômbia (32º lugar), Bolívia (60º) e Venezuela (67º). Essa dimensão avalia as diferenças entre homens e mulheres em aspectos como inserção no mercado de trabalho, renda e equidade salarial.

A exclusão financeira pode ser vista não apenas como um problema individual, mas também como um desafio para o desenvolvimento econômico de comunidades inteiras. Sem acesso a serviços financeiros, as pessoas tendem a recorrer a alternativas informais, que geralmente são mais caras e menos seguras. Isso pode incluir agiotagem, esquemas financeiros informais e poupanças feitas em casa, além de jogos online, os quais não oferecem a mesma proteção e estabilidade que os serviços financeiros regulamentados. Isso pode ser observado no gráfico abaixo de uma pesquisa realizada pelo Banco Central do Brasil:

GRÁFICO 4.8.1.1 – O QUE VOCÊ FEZ PARA PAGAR AS CONTAS, COBRIR TODAS AS DESPESAS?

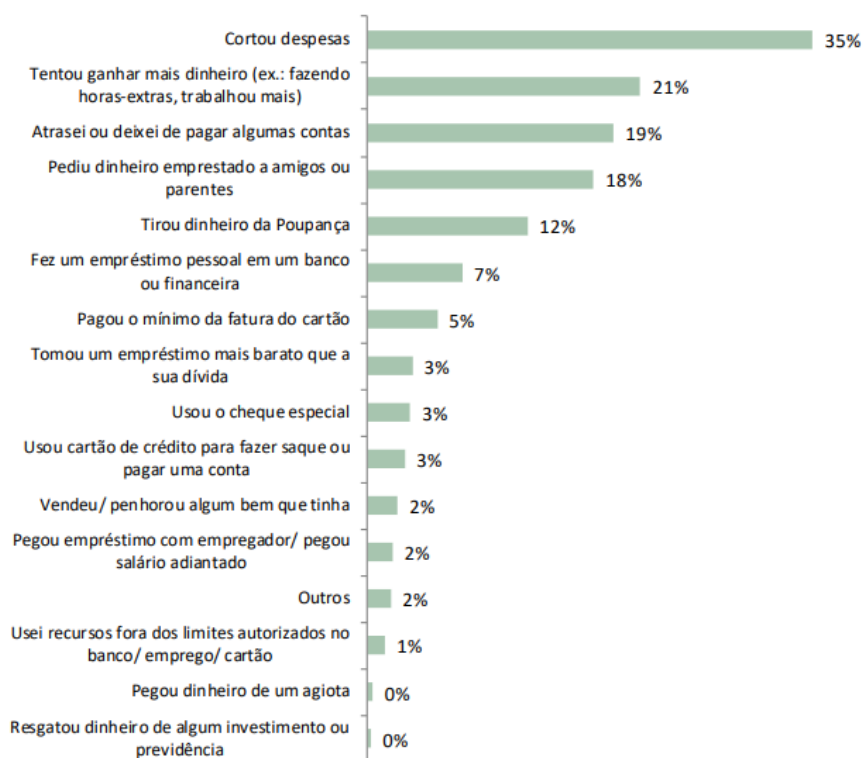


Figura 3 - Gastos além do orçamento (Banco Central do Brasil, 2017)

De acordo com a Nota Técnica 513/2024-BCB/SECRE do Banco Central do Brasil (2024), estima-se que, em agosto de 2024, 5 milhões de pessoas de famílias beneficiárias do Programa Bolsa Família utilizaram a plataforma Pix para enviar R\$3 bilhões a empresas de apostas. A mediana dos valores apostados foi de R\$100 por pessoa. Entre esses apostadores, 4 milhões (70%) são chefes de família, responsáveis por receber o benefício, e movimentaram R\$2 bilhões (67%) por meio do Pix para apostas.

O tema da educação financeira, conforme destacado pelo Banco Central do Brasil (BACEN, 2017), tem ganhado relevância na agenda política global, uma vez que desempenha um papel fundamental no empoderamento financeiro dos cidadãos. Essa conscientização financeira, por sua vez, contribui positivamente para a estabilidade do sistema financeiro e da economia. A falta de educação financeira adequada leva à má gestão de recursos pessoais, resultando em endividamento excessivo, baixa capacidade de poupança e vulnerabilidade financeira, afetando tanto o indivíduo quanto a estabilidade macroeconômica.

De acordo com Crocco et. al. (2013), o combate à exclusão financeira tem sido alvo de medidas de intervenção governamental, que podem seguir duas abordagens. Em primeiro lugar, o governo atua como facilitador, promovendo ações para combater a exclusão financeira, verificando se os produtos bancários atingem esse objetivo e pressionando por regulamentações. Na segunda, o governo atua como legislador, impondo a oferta de serviços bancários universais ou removendo obstáculos que reforçam a exclusão financeira, como restrições à mobilidade de capital. Além disso, incentivos diretos, como isenções fiscais e garantias de crédito, também podem ser usados para motivar a inclusão financeira.

Nos países em desenvolvimento, onde a desbancarização é mais equipada, a inclusão financeira tem sido promovida através do design de produtos e serviços digitais. Vargas e Santos (2021) confirmam o sucesso das plataformas digitais na redução da desigualdade, ressaltando que a inclusão financeira, ao lado da inclusão digital, tem permitido a entrada de novos atores, impactando positivamente o âmbito econômico e social. O Brasil, ocupando o terceiro lugar entre os países do BRICS em termos de inclusão financeira, destaca-se na América Latina e deve considerar estratégias de longo prazo, como a alfabetização financeira e a colaboração com a sociedade civil.

Em suma, a discriminação e a exclusão financeira são barreiras significativas ao desenvolvimento econômico e à igualdade social. Abordá-las requer um esforço conjunto de governos, instituições financeiras e sociedade civil. Ao promover uma maior inclusão financeira, é possível criar um ambiente onde todos tenham a oportunidade de prosperar e contribuir para o bem-estar econômico da sociedade como um todo.

3.2.2 Sustentabilidade Financeira e Responsabilidade Ética

Os Objetivos de Desenvolvimento Sustentável (ODS), estabelecidos pela ONU, desempenham um papel crucial na mitigação da exclusão financeira ao promover a inclusão econômica global, com foco em reduzir desigualdades e garantir que todos tenham acesso a serviços financeiros acessíveis e equitativos, fundamentais para o desenvolvimento sustentável. Constituem uma agenda global adotada pelas Nações Unidas em 2015, visando erradicar a pobreza, proteger o planeta e garantir paz e prosperidade para todos até 2030. São 17 objetivos interconectados, conforme citado por Roma (2019) que, em seu conjunto, almejavam fazer com que o mundo progredisse rapidamente rumo à eliminação da extrema pobreza e da fome do planeta, fatores que afetam especialmente as populações mais pobres, dos países menos desenvolvidos.

De acordo com as Nações Unidas Brasil (2015), os Objetivos de Desenvolvimento Sustentável representam um apelo global para ações que visam erradicar a pobreza, proteger o meio ambiente e o clima, e garantir que todas as pessoas possam desfrutar de paz e prosperidade. Esses objetivos são o foco dos esforços das Nações Unidas para alcançar a Agenda 2030 no Brasil.

Para criar um campo de jogo verdadeiramente equitativo, pode ser necessário um enfoque integrado e holístico. Investir na educação financeira desde cedo pode capacitar indivíduos a fazerem escolhas mais informadas. Reformas econômicas que facilitem o acesso ao capital e incentivem o empreendedorismo também podem ser essenciais.

Além disso, políticas rígidas contra discriminação de crédito precisam ser implementadas e fiscalizadas, garantindo que todos tenham oportunidades justas de prosperar no setor financeiro. A transição para um sistema econômico mais ético e responsável pode não ser apenas desejável, mas essencial para a construção de um futuro que respeite tanto os direitos humanos quanto o meio ambiente, assegurando que as comunidades vulneráveis sejam protegidas e que o planeta possa prosperar.

MITIGAÇÃO DO VIÉS ALGORÍTMICO

4.1. Técnicas de Mitigação

Para mitigar riscos, instituições financeiras podem adotar uma série de técnicas. A implementação de sistemas de alerta precoce pode permitir a detecção de anomalias e padrões irregulares em tempo real. O uso de algoritmos de *machine learning* para análise de crédito pode melhorar a avaliação de solvência dos clientes, reduzindo a inadimplência. A criptografia e outras medidas de segurança cibernética são consideradas cruciais para proteger dados sensíveis contra acessos não autorizados e ataques cibernéticos.

Um dos métodos mais comuns para controle do risco é o sistema de score. Esse sistema funciona através da avaliação de características de um novo cliente, atribuindo um valor específico a cada uma delas. Esses dados são, então, utilizados para calcular um score. Com o score gerado, decide-se se o crédito será concedido ou não. Para essa decisão, o score é comparado a um limite pré-determinado, conhecido como valor de corte. Os principais desafios enfrentados pelos profissionais da área estão relacionados à definição desse valor (GUIMARÃES; CHAVES NETO, 2002).

A adoção de novas soluções tecnológicas no setor financeiro abre a possibilidade de antecipar riscos regulatórios, permitindo uma avaliação mais precisa

das potenciais ameaças à estabilidade financeira e ao equilíbrio sistêmico. A supervisão constante e o escopo limitado dos testes ajudam a mitigar esses riscos, mas o regulador deve estar atento às rápidas transformações do mercado, ajustando o arcabouço normativo de forma ágil e eficaz para garantir que os comportamentos inovadores não comprometam a segurança do sistema financeiro como um todo (UEDA, 2023).

Lee et al. (2019) sugerem que, embora as leis não necessariamente mitiguem ou resolvam preconceitos implícitos ou inconscientes que podem ser incorporados aos algoritmos, as empresas e outros operadores devem se proteger contra a violação dessas proteções estatutárias no design de algoritmos. Além disso, é importante mitigar a preocupação implícita de impedir que a discriminação passada continue.

Outra abordagem eficaz é a análise de sentimento, que avalia o humor do mercado através de mídias sociais e notícias financeiras, permitindo que investidores ajustem suas estratégias rapidamente. Com o crescimento das redes sociais, como o X® (antigo *Twitter*®), essas plataformas se tornaram essenciais na previsão das movimentações do mercado financeiro, uma vez que postagens sobre economia podem influenciar decisões de investidores, criando um efeito borboleta ao disseminar rapidamente informações (YAN, 2023, apud ARAUJO; FERNANDES, 2021; REED, 2016). Segundo Yan (2023), algoritmos de Aprendizado de Máquina e Aprendizado Profundo são eficazes na previsão de ações que utilizaram modelos de linguagem e redes neurais para analisar sentimentos em postagens, obtendo resultados promissores.

Portanto, a análise de dados e a detecção de padrões no setor financeiro não apenas fomentam uma compreensão mais profunda do mercado, mas também permitem a criação de estratégias robustas para mitigação de riscos. Quando um analista de crédito dispõe de um sistema de reconhecimento de padrões e classificação que fornece uma previsão sobre a probabilidade de inadimplência de um potencial cliente, o processo de decisão quanto à concessão de crédito se torna significativamente mais simples. Dessa forma, o analista pode basear suas escolhas em dados quantitativos, ao invés de opiniões subjetivas, aumentando sua confiança nas decisões tomadas (GUIMARÃES; CHAVES NETO, 2002).

Integrar uma abordagem de direitos humanos na construção e implementação de sistemas financeiros e tecnologias é imprescindível para fomentar o desenvolvimento sustentável e a justiça social. A mitigação do viés algorítmico torna-

se, portanto, uma componente essencial nesse cenário, garantindo que as soluções financeiras tecnológicas não apenas evitem discriminações, mas também promovam a equidade.

Yoshinaga e Castro (2023) apontam que a crescente adoção da IA nas finanças levanta considerações regulatórias e éticas. Entre os desafios a serem enfrentados estão a privacidade de dados, a transparência algorítmica e os possíveis vieses em algoritmos de IA. Órgãos reguladores estão trabalhando para estabelecer estruturas que garantam o uso responsável e ético da IA na indústria financeira.

Segundo Yoshinaga & Castro (2023), as finanças têm se beneficiado da inteligência artificial (IA), com algoritmos de aprendizado de máquina processando dados históricos para identificar padrões que analistas humanos podem perder. Esses algoritmos ajudam a prever movimentos de mercado e a otimizar a alocação de ativos, aumentando a precisão das previsões e reduzindo o viés humano nas decisões. Contudo, os investidores devem estar cientes de riscos, como a dependência de dados históricos e a possibilidade de ajuste excessivo (*overfitting*), tornando essencial um desenvolvimento e monitoramento adequado dos modelos de IA.

Ainda, segundo Yoshinaga & Castro (2023), os algoritmos de IA por serem fortemente dependentes de dados para treinamento e tomada de decisões podem resultar em decisões injustas ou discriminatórias, perpetuando vieses sociais, econômicas ou culturais. Portanto, é essencial garantir que os dados de treinamento sejam de alta qualidade e variados, além de implementar medidas para mitigar esses vieses, a fim de promover a justiça e a imparcialidade e evitar danos potenciais. Experimentos mostram que o uso de IA na seleção de currículos pode manter padrões de contratação não diversos devido à qualidade da base de dados utilizada para o treinamento do algoritmo.

A integração de princípios de direitos humanos na tecnologia financeira pode ser vista não apenas como uma reflexão ética, mas como uma necessidade prática para construir uma economia mais justa e inclusiva. Uma abordagem que combine inovação tecnológica com justiça social tem o potencial de não só atingir, mas também superar, os objetivos de desenvolvimento sustentável, criando um futuro em que o progresso econômico e a igualdade caminhem lado a lado.

4.2. Governança e Compliance

Conforme Valle e Santos (2019), o conceito de governança corporativa vai além da simples distribuição de poder, envolvendo um modelo de atuação empresarial

que integra princípios como transparência, objetividade e equidade no tratamento de seus sócios, além de reforçar a responsabilidade perante terceiros que fornecem recursos. Em complemento, o conceito de *compliance* é entendido como “um conjunto de regras, padrões, procedimentos éticos e legais, que, uma vez definido e implantado, será a linha mestra que orientará o comportamento da instituição no mercado em que atua, bem como a atitude dos seus funcionários”, reforçando o compromisso com a conformidade legal e a ética nos negócios.

Conforme Trivelato et. al. (2018), as empresas brasileiras ainda não aderiram completamente aos preceitos da governança corporativa, o que abre espaço para a disseminação de seus valores, como *fairness* (equidade), *disclosure* (transparência), *accountability* (responsabilidade) e *compliance* (conformidade). A governança corporativa, como um sistema que guia as organizações, visa agregar valor ao negócio, proporcionando benefícios e vantagens competitivas. Além disso, o *compliance* desempenha um papel crucial ao evitar e/ou mitigar riscos e problemas derivados da má governança, que pode ir contra as questões legais e éticas da organização.

Nos últimos anos, a digitalização e a inovação tecnológica trouxeram novos desafios para a governança e o *compliance* no setor financeiro. O surgimento de *fintechs*, *big data*, *blockchain* e inteligência artificial altera significativamente o panorama regulatório, exigindo adaptações rápidas e eficazes para manter a conformidade. Além disso, a crescente demanda por transparência nos procedimentos financeiros e o aumento de crimes cibernéticos intensificam a necessidade de estruturas robustas de governança e *compliance*.

Lumango (2020) observa que as instituições financeiras agora precisam desenvolver novas plataformas, indo além de simples atualizações das existentes. As *startups* financeiras, seja em parceria com incumbentes ou explorando novos segmentos de mercado de forma independente, também enfrentam o desafio de dimensionar seus negócios enquanto lidam com uma regulamentação e governança mais rigorosas. Esse novo paradigma de concorrência, trazido por *startups* de fora do sistema financeiro, introduziu uma nova onda de inovações e desenvolvimento no setor, principalmente através da incorporação de tecnologia nos processos de negócio.

De acordo com De Oliveira Filho e Da Cruz David (2022), a participação das empresas no sistema de implementação de Direitos Humanos não é mais uma opção

ocasionalmente vantajosa, mas uma exigência do estágio atual de evolução histórica, tanto em nível internacional quanto nacional. Isso reforça a responsabilidade corporativa em alinhar suas práticas com princípios que promovam o respeito aos Direitos Humanos, transcendendo o mero cumprimento de normas regulatórias.

Portanto, a governança corporativa e o *compliance* antidiscriminatório podem ser fundamentais para garantir que as empresas atuem de forma ética e transparente, prevenindo violações de direitos e promovendo a equidade dentro e fora de suas operações. Essas práticas permitem uma abordagem proativa na mitigação de riscos, como o viés algorítmico, ao garantir que as decisões sejam pautadas em padrões justos e imparciais, respeitando a dignidade humana.

Em suma, a integração de governança e *compliance* com uma perspectiva de Direitos Humanos pode não apenas fortalecer a confiança social nas instituições, mas também promover um ambiente empresarial mais justo e sustentável.

4.2.1 Política de Responsabilidade Social, Ambiental e Climática (PRSAC)

A Política de Responsabilidade Social, Ambiental e Climática (PRSAC) é, conforme Banco Central do Brasil (2021), um conjunto de diretrizes e procedimentos que empresas e instituições adotam visando promover práticas sustentáveis e equitativas em suas operações. A PRSAC busca conciliar crescimento econômico com responsabilidade social, respeito ao meio ambiente e atenção às mudanças climáticas. A implementação dessa política envolve a combinação de ações de governança corporativa e *compliance* para assegurar que as atividades empresariais estejam alinhadas com valores éticos e sustentáveis.

No contexto das instituições financeiras, a PRSAC pode desempenhar um papel crucial na mitigação do viés algorítmico. Algoritmos são ferramentas poderosas, mas podem ser influenciados por dados históricos que refletem desigualdades e preconceitos sociais. Com uma política de responsabilidade social, essas instituições são incentivadas a revisar constantemente os dados utilizados para treinar algoritmos e a implementar medidas de transparência e equidade. Isso pode ajudar a garantir que as decisões automatizadas, como concessão de créditos ou seguros, sejam justas e não perpetuem discriminações.

A PRSAC também pode se alinhar com os princípios de direitos humanos ao promover práticas que respeitam a dignidade e igualdade de todos os indivíduos. Empresas que aderem a essa política tendem a se engajar em atividades como auditorias de direitos humanos, iniciativas de diversidade e inclusão e parcerias com organizações não-governamentais. Essas ações não só podem prevenir violações de direitos humanos, mas também criar um ambiente mais inclusivo e justo para colaboradores e comunidades envolvidas.

No combate à desigualdade social, a PRSAC pode fomentar a implementação de programas de responsabilidade social corporativa, como educação financeira para comunidades carentes, investimentos em infraestrutura sustentável e promoção de condições de trabalho justas. Essas iniciativas ajudam a reduzir a disparidade econômica e promovem o desenvolvimento sustentável em comunidades vulneráveis.

A governança e o *compliance* são pilares fundamentais para a efetividade da PRSAC. Uma governança corporativa sólida pode garantir que as decisões estratégicas da empresa sejam tomadas com responsabilidade e transparência. Já o *compliance* assegura que todas as operações estejam em conformidade com leis e regulamentos relevantes, evitando práticas nocivas e garantindo a integridade da empresa.

Em resumo, a PRSAC pode ser uma ferramenta vital para instituições financeiras e outras empresas que desejam operar de maneira ética e sustentável. Ao abordar questões relacionadas ao viés algorítmico, direitos humanos e desigualdade social através de mecanismos robustos de governança e *compliance*, essa política não só promove um ambiente corporativo mais justo, mas também contribui para um impacto positivo duradouro na sociedade.

REGULAÇÃO E POLÍTICAS PÚBLICAS

5.1. Regulações Existentes

A ascensão dos algoritmos nas decisões de várias áreas da vida moderna, como contratação de funcionários e avaliação de crédito, traz à tona uma série de preocupações éticas e sociais, principalmente no que diz respeito a vieses e discriminação algorítmica. Este fenômeno se torna ainda mais preocupante ao ser analisado sob a ótica dos direitos humanos, conforme definido na Declaração Universal dos Direitos Humanos (1948) e em outros instrumentos internacionais.

No contexto brasileiro, a Constituição Federal de 1988 consagra diversos direitos humanos, como a dignidade da pessoa humana, igualdade e não discriminação. Nesse sentido, o uso de algoritmos que possam violar esses direitos desafia a legislação existente, exigindo uma regulamentação adequada que evite a perpetuação de práticas discriminatórias.

A Lei Geral de Proteção de Dados (LGPD) - Lei nº 13.709/2018, é um marco importante nessa questão. Além de regular o tratamento de dados pessoais, a LGPD estabelece direitos fundamentais que protegem os cidadãos de decisões automatizadas potencialmente discriminatórias. O Artigo 20, por exemplo, garante ao titular o direito de contestar decisões automatizadas, assegurando que o uso de dados respeite a dignidade e os direitos humanos. Esse dispositivo é essencial para mitigar o impacto negativo de algoritmos que possam reproduzir preconceitos implícitos nos dados utilizados para treiná-los.

Outra legislação relevante é o Estatuto da Igualdade Racial (Lei nº 12.288/2010), que busca combater a discriminação racial e promover a igualdade de oportunidades. A implementação de algoritmos em contextos que afetam grupos racialmente discriminados deve ser cuidadosamente monitorada para evitar a perpetuação de preconceitos. Além disso, o Plano Nacional de Direitos Humanos (PNDH) estabelece diretrizes gerais para a promoção e proteção dos direitos humanos no Brasil. Embora não se refira diretamente ao viés algorítmico, sua aplicação pode exigir medidas que previnam a discriminação e promovam a equidade no uso de tecnologias emergentes.

No contexto global, a Declaração Universal dos Direitos Humanos (1948) estabelece que todos os seres humanos têm direito à igualdade e à não discriminação, princípios que podem ser desafiados pelo uso de tecnologias algorítmicas tendenciosas. O Pacto Internacional sobre Direitos Civis e Políticos (1966) reforça esses direitos, assegurando a igualdade perante a lei e a proteção contra a

discriminação. Portanto, os mecanismos de Inteligência Artificial (IA) e algoritmos devem ser avaliados em relação a esses direitos internacionais.

A União Europeia tem avançado com legislações como o Regulamento Geral sobre a Proteção de Dados (GDPR), que protege dados pessoais e reforça os direitos dos indivíduos em relação a decisões automatizadas. A transparência e a capacidade de contestar essas decisões automatizadas são alinhadas com os princípios de direitos humanos. Adicionalmente, a Proposta de Regulação da Inteligência Artificial (AI Act) visa garantir que os sistemas de IA respeitem direitos fundamentais e previnam discriminações, especialmente em sistemas de alto risco que possam afetar os direitos dos cidadãos.

O Relatório de 2021 do Alto Comissariado da ONU sobre IA e Direitos Humanos enfatiza a importância de garantir que as tecnologias de IA respeitem os direitos humanos. O documento recomenda que os Estados adotem legislações específicas para evitar discriminações e proteger os direitos humanos na era da automação.

Assim, a interseção entre legislações de direitos humanos e viés algorítmico é fundamental para assegurar que as tecnologias emergentes sejam desenvolvidas e usadas de maneira a promover a dignidade humana. À medida que algoritmos se tornam mais influentes na tomada de decisões que impactam a vida das pessoas, a necessidade de estruturas legais robustas que protejam contra discriminações e garantam transparência é cada vez mais urgente. Portanto, é imperativo que legislações tanto no Brasil quanto globalmente evoluam para enfrentar esses desafios contemporâneos, garantindo que o progresso tecnológico não ocorra à custa dos direitos humanos fundamentais.

Observando as declarações e resoluções internacionais, como a Declaração das Nações Unidas sobre o Direito ao Desenvolvimento (1986) e a Declaração das Nações Unidas sobre os Direitos dos Povos Indígenas (2007), torna-se claro que o desenvolvimento e uso de tecnologias algorítmicas devem promover justiça social e inclusão. A Recomendação da UNESCO sobre a Ética da Inteligência Artificial (2022) também enfatiza a necessidade de que estas tecnologias sejam desenvolvidas e usadas de maneira justa e inclusiva.

Finalmente, a implementação de regulamentações eficazes em relação a vieses e discriminação algorítmica pode também depender da educação e capacitação. Desenvolvedores de IA e tomadores de decisão precisam ser

conscientizados sobre os impactos de vieses algorítmicos e equipados com ferramentas para mitigá-los. Além disso, a sociedade civil deve ser engajada no processo regulatório para garantir que as soluções adotadas reflitam uma ampla gama de perspectivas e necessidades.

5.2. Propostas de Novas Políticas

A intersecção entre governança, *compliance*, PRSAC (Prevenção à Responsabilidade Social, Ambiental e *Compliance*) e direitos humanos pode oferecer uma base sólida para a formulação de políticas públicas que mitiguem o viés algorítmico e melhorem os direitos humanos no setor financeiro. Com a crescente dependência de algoritmos e inteligência artificial, emergem preocupações significativas sobre discriminação involuntária, transparência e equidade.

O problema identificado é a falta de garantia de que algoritmos empregados em processos financeiros sejam justos e equitativos. Uma política pública deve estabelecer diretrizes claras para garantir a transparência algorítmica. Bancos e outras instituições financeiras poderiam ser obrigados a divulgar os critérios usados em seus algoritmos de avaliação de crédito e detecção de fraudes. Isso incluiria a obrigatoriedade de auditorias externas regulares para revisar e validar esses algoritmos, garantindo que não perpetuem vieses discriminatórios.

Implementar programas de formação e capacitação contínua para desenvolvedores e funcionários de instituições financeiras sobre vieses algorítmicos e direitos humanos pode ser essencial. A educação desses profissionais é crucial para a criação de algoritmos que respeitem a diversidade e a inclusão. Treinamentos específicos em ética digital e análise crítica dos dados empregados podem ajudar a mitigar preconceitos inconscientes.

Estabelecer comissões independentes para a supervisão de práticas algorítmicas no setor financeiro, compostas por especialistas em ética, tecnologia e direitos humanos, pode ser uma medida eficaz. Essas comissões seriam responsáveis por monitorar e avaliar o impacto das tecnologias

adotadas, garantindo que as práticas empresariais estejam alinhadas com os princípios de direitos humanos e não discriminem qualquer grupo social.

Desenvolver políticas que exijam a utilização de conjuntos de dados diversificados e representativos na criação e treinamento de algoritmos financeiros pode ajudar a criar sistemas mais equitativos. Dados enviesados muitas vezes resultam em decisões que prejudicam grupos minoritários. Promover a inclusão de variáveis que representem de maneira justa diferentes segmentos sociais pode ser uma solução.

Implementar mecanismos robustos de feedback e reavaliação contínua dos algoritmos financeiros pode ser fundamental. Usuários afetados por decisões algorítmicas, como a negação de crédito, devem ter o direito de contestar essas decisões e fornecer feedback que possa levar à reavaliação dos algoritmos. Um sistema de coleta de feedback bem estruturado ajudaria a identificar e corrigir rapidamente qualquer viés que venha a surgir.

Integrar as melhores práticas de *compliance* e governança especificamente adaptadas para o contexto de algoritmos financeiros pode incluir a criação de políticas claras sobre a responsabilidade legal das instituições financeiras em casos de discriminação algorítmica, além de incentivar lideranças a adotar práticas éticas na gestão de tecnologia.

Oferecer incentivos fiscais e financeiros a instituições financeiras que adotem práticas inovadoras de mitigação de viés e que demonstrem um compromisso ativo com os direitos humanos pode ser uma estratégia eficaz. Programas de certificação de práticas *fintechs* inclusivas podem ser desenvolvidos para reconhecer e promover instituições com alto desempenho nesses critérios.

Ao implementar essas propostas, pode ser possível criar um ambiente financeiro mais justo e inclusivo, onde a tecnologia apoia os direitos humanos e a equidade. Essas políticas públicas não apenas mitigarão o viés algorítmico, mas também promoverão uma cultura de transparência e responsabilidade social no setor financeiro.

5.2.1. Diretrizes para Empresas e Desenvolvedores

A mitigação do viés algorítmico pode ser fundamental para garantir a justiça e equidade em sistemas de IA, como já abordado neste trabalho. Empresas e desenvolvedores devem adotar práticas que promovam diversidade e inclusão desde

o design dos algoritmos, considerando as diferentes necessidades dos usuários. Essa responsabilidade envolve aspectos como privacidade, transparência, segurança e educação, assegurando que a tecnologia contribua positivamente para a sociedade.

A transparência é essencial, permitindo que métricas de equidade sejam comunicadas aos usuários e possibilitando auditorias independentes que identifiquem possíveis injustiças. A diversidade nas equipes de desenvolvimento, com pessoas de diferentes contextos culturais e sociais, também pode ser crucial para mitigar vieses. Auditorias frequentes e o uso de ferramentas de monitoramento contínuo ajudam a identificar e corrigir desvios nos algoritmos, assegurando que funcionem de forma justa e ética.

Além disso, a explicabilidade dos modelos de IA é um pilar importante, aumentando a confiança dos usuários e facilitando a regulamentação desses sistemas. A coleta e o pré-processamento rigoroso de dados, juntamente com o treinamento contínuo das equipes em diversidade e mitigação de viés, são essenciais para evitar injustiças nos algoritmos.

Investir em algoritmos que promovam explicitamente a equidade pode ser vital, especialmente em áreas como saúde, educação e justiça. Técnicas como regularização equitativa e reponderação podem ajudar a otimizar tanto a precisão quanto a justiça dos modelos. A interação com *stakeholders* também é importante, permitindo ajustes e promovendo a confiança no sistema.

DISCUSSÃO

6.1. Futuras Tendências e Pesquisas

As futuras tendências e pesquisas no campo do viés algorítmico, combinadas com preocupações de direitos humanos e finanças sustentáveis, podem visar promover maior equidade e inclusão na era digital. Diversas áreas mostram-se promissoras para investigações futuras:

Primeiramente, o desenvolvimento de algoritmos justos e transparentes é uma tendência crucial. A pesquisa está focada em criar algoritmos que não apenas sejam precisos, mas também equânimes e transparentes. Isso inclui a criação de técnicas para detectar e mitigar vieses em sistemas de IA, além de métodos para auditar esses algoritmos, garantindo que não perpetuem discriminações.

Outra área vital é a integração de direitos humanos nas políticas de IA. As futuras pesquisas deverão se concentrar na formulação de políticas públicas e diretrizes que incorporem explicitamente os direitos humanos à governança das tecnologias emergentes. Isso envolve o desenvolvimento de *frameworks* regulatórios que assegurem a equidade e protejam os direitos dos indivíduos contra decisões automatizadas.

Também é fundamental investigar o impacto do viés algorítmico em diversos setores além do financeiro, como saúde, educação e justiça. Estudos futuros podem explorar como o viés em sistemas de IA influencia a distribuição de recursos e oportunidades em diferentes contextos sociais e econômicos.

No campo das finanças, a pesquisa continuará explorando como as finanças sustentáveis podem ser aprimoradas para promover maior inclusão financeira. Isso inclui analisar como as tecnologias financeiras podem ser usadas para reduzir desigualdades e melhorar o acesso a serviços financeiros para populações marginalizadas, sem comprometer justiça e equidade.

A educação e capacitação em IA destacam-se como outra área de pesquisa essencial. Formar profissionais em IA e dados com ênfase em ética e direitos humanos será crucial. Pesquisas futuras podem focar em como melhorar a educação e capacitação, garantindo que desenvolvedores e gestores de IA estejam conscientes das implicações éticas e sociais de suas criações.

Além disso, é importante monitorar os impactos de novas tecnologias emergentes. A introdução de inovações como IA explicativa e IA responsável deve ser acompanhada de perto. As tendências futuras podem investigar como essas

tecnologias podem tornar os sistemas de IA mais compreensíveis e auditáveis, reduzindo o viés e promovendo a justiça.

O desenvolvimento de mecanismos robustos para o monitoramento e avaliação contínua dos sistemas de IA também é uma área em crescimento. Isso envolve criar ferramentas e métricas que permitam acompanhar o desempenho dos algoritmos em tempo real e identificar possíveis desvios éticos e viés.

Por fim, a colaboração interdisciplinar e global será fundamental para enfrentar os desafios do viés algorítmico. Estudos futuros podem explorar como parcerias globais e interdisciplinares entre pesquisadores, formuladores de políticas, empresas e organizações da sociedade civil podem contribuir para soluções mais eficazes e abrangentes.

Essas áreas de pesquisa promissoras são vitais para assegurar que o avanço das tecnologias digitais seja acompanhado por um comprometimento contínuo com a equidade, inclusão e respeito aos direitos humanos.

CONCLUSÃO

Concluir este estudo sobre os efeitos do viés algorítmico nos direitos humanos no setor financeiro pode permitir uma compreensão mais profunda e complexa desta questão. Através da análise detalhada, observa-se que, apesar do poder transformador da inteligência artificial e do aprendizado de máquina, há consideráveis riscos associados, particularmente em termos de discriminação e exclusão financeira.

A fundamentação teórica proporciona uma compreensão de conceitos fundamentais como inteligência artificial, tipos de algoritmos e viés algorítmico. O viés pode surgir de várias fontes, incluindo dados inadequados, processos de seleção imperfeitos e preconceitos humanos, que são inadvertidamente incorporados nos algoritmos. Problemas cruciais, como a discriminação sistemática e a desigualdade de oportunidades que emergem das práticas algorítmicas tendenciosas, representam sérias violações dos direitos humanos e dos objetivos de desenvolvimento sustentável (ODS).

A pesquisa discute métodos eficazes para identificar e mitigar o viés algorítmico, enfatizando a importância de algoritmos de equidade, governança robusta e *compliance*. As políticas de responsabilidade social, ambiental e climática (PRSAC) são identificadas como instrumentos essenciais para garantir operações mais justas e inclusivas.

Por fim, uma avaliação das regulações existentes e propostas de novas políticas revela um campo em evolução, onde a adaptação rápida é necessária para enfrentar desafios emergentes. Recomendações específicas para empresas e desenvolvedores destacam a necessidade de uma abordagem proativa e integrada para salvaguardar os direitos humanos no contexto da transformação digital.

Ao avançar, é crucial abordar os desafios e limitações identificados e continuar explorando novas pesquisas e tendências futuras. Somente através de um esforço contínuo e colaborativo será possível mitigar os riscos e maximizar os benefícios da tecnologia para todos.

Conclui-se que combater o viés algorítmico é uma tarefa complexa, exigindo uma abordagem integrada que envolva aspectos técnicos, éticos e regulatórios. A colaboração entre tecnólogos, legisladores e a sociedade civil é essencial para criar um ambiente financeiro que respeite os direitos humanos e promova a justiça social. Este trabalho contribui para a discussão contínua e sugere áreas para futuras

pesquisas, incluindo o desenvolvimento de novos métodos de mitigação e a análise do impacto das políticas públicas implementadas.

Portanto, é crucial continuar investindo em soluções inovadoras e eficazes para enfrentar os desafios do viés algorítmico, garantindo que as tecnologias emergentes cumpram seu potencial de forma ética e inclusiva.

REFERÊNCIAS

ALMEIDA, Adriano; CARVALHO, Felipe; MENINO, Felipe. **Introdução ao machine learning**. Disponível em: <https://dataat.github.io/introducao-ao-machine-learning/index.html>. Acesso em: 02 ago. 2024.

BANCO CENTRAL DO BRASIL. **Estudos sobre educação, proteção e inclusão**. 2017. Disponível em: https://www.bcb.gov.br/nor/reincfin/serie_cidadania_financeira_pesquisa_infe_br_%200443_2017.pdf. Acesso em: 02 ago. 2024.

BANCO CENTRAL DO BRASIL. **Estudo sobre o perfil de apostadores no Brasil**. Brasília, 2024. Disponível em: <https://static.poder360.com.br/202/09/Husa--apostas-bc--24set2024.pdf>. Acesso em: 25 set. 2024.

BANCO CENTRAL DO BRASIL. **Resolução CMN nº 4.945, de 15 de setembro de 2021**. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20CMN&numero=4945>. Acesso em: 02 ago. 2024.

BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. **Big data's disparate impact**. SSRN. 2014. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899. Acesso em: 02 ago. 2024.

BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. **Fairness and machine learning**. 2023. Disponível em: <https://fairmlbook.org/>. Acesso em: 02 ago. 2024.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Diário Oficial da União, Brasília, DF, 5 ago. 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 03 ago. 2024.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 03 ago. 2024.

BRASIL. **Lei nº 12.288, de 20 de julho de 2010**. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/lei/l12288.htm. Acesso em: 03 ago. 2024.

BRASIL. **Decreto nº 7.037, de 21 de dezembro de 2009**. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/decreto/d7037.htm. Acesso em: 03 ago. 2024.

CARNEIRO, Fábio Lacerda. **Modelagem de risco de crédito de portfólio: implicações para a regulamentação sobre requerimento de capital de instituições financeiras**. 2002. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2002.

CHATGPT. **Definições de termos relacionados a direitos humanos e inteligência artificial**. Consultas realizadas em 06 de outubro de 2024. Disponível em: <https://www.openai.com/chatgpt>. Acesso em: 06 out. 2024.

COZMAN, Fabio Gagliardi; KAUFMAN, Dora. **Viés no aprendizado de máquina em sistemas de inteligência artificial: a diversidade de origens e os caminhos de mitigação**. *Revista USP*, n. 135, p. 195-210, 2022.

CROCCO, Marco Aurelio; SANTOS, Fabiana; FIGUEIREDO, Ana. **Exclusão financeira no Brasil: uma análise regional exploratória**. *Brazilian Journal of Political Economy*, v. 33, p. 505-526, 2013.

DA ROCHA, Cláudio Jannotti; PORTO, Lorena Vasconcelos; ABAURRE, Helena Emerick. **Discriminação algorítmica no trabalho digital**. *Revista de Direitos Humanos e Desenvolvimento Social*, v. 1, p. 1-21, 2020.

DA SILVA CARAFFINI, Josiane Piva Testolin; DE SOUZA, Romina Batista de Lucena; BEHR, Ariel. **Transformação digital e desempenho no setor bancário**. In: Congresso Transformação Digital 2018. 2018.

DE JESUS, Everaldo Antônio et al. **Impacto da inteligência artificial na sociedade contemporânea**. *Revista Amor Mundi*, v. 5, n. 5, p. 43-58, 2024.

DE PÁDUA, Sérgio R.; LORENZETTO, Bruno M. **O direito fundamental à explicabilidade da inteligência artificial utilizada em decisões estatais**. *Revista da AGU*. 2024. Disponível em: <https://revistaagu.agu.gov.br/index.php/AGU/article/view/3480>. Acesso em: 02 ago. 2024.

DE OLIVEIRA FILHO, João Glicério; DA CRUZ DAVID, Roberto. **Governança corporativa e compliance antidiscriminatório no Brasil: um estudo sobre empresa e direitos humanos**. *Diálogos Possíveis*, v. 21, n. 2, 2022.

DOS SANTOS, Fábio Oliveira **Uso de algoritmos de aprendizado de máquina na resolução de falhas em redes móveis** / Fábio Oliveira Dos santos ; Natalia Castro Fernandes, orientadora. Niterói, 2019. 100 f. : il.

FEBRABAN. **Pesquisa Febraban de tecnologia bancária 2024**. Disponível em: <https://cmsarquivos.febraban.org.br/Arquivos/documentos/PDF/Pesquisa%20Febraban%20de%20Tecnologia%20Banc%C3%A1ria%202024.pdf>. Acesso em: 02 ago. 2024.

GARCIA, Ana Cristina Bicharra; GARCIA, Marcio Gomes Pinto; RIGOBON, Roberto. **Algorithmic discrimination in the credit domain: what do we know about it?**. *AI & Society*, 2023. Disponível em: <https://doi.org/10.1007/s00146-023-01676-3>. Acesso em: 05 ago. 2024.

GARCÍA et al. **Desigualdade vs igualdade vs equidade vs justiça: uma explicação visual**. *Journal of Computational Intelligence*, v. 37, n. 4, p. 123-145, 2023. Disponível em: <https://link.springer.com/article/10.1007/s00146-023-01676-3/figures/1>. Acesso em: 14 set. 2024.

GHISLENI, Eduardo Steffenello. **Vigilância na sociedade em rede: a coleta de dados pessoais na internet e suas implicações ao direito à privacidade**. 2015. Dissertação (Mestrado) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2015.

GOULART, Guilherme. **O impacto das novas tecnologias nos direitos humanos e fundamentais: o acesso à internet e a liberdade de expressão**. *Revista Brasileira de Políticas Públicas*, v. 1, n. 10, 2012. ISSN 2316-3054.

GONZAGA, Carolina de Moura et al. **Os impactos causados pela pandemia do Coronavírus: COVID-19 no ramo empresarial**. 2023.

GONZALES, Julius Tan. Implications of AI innovation on economic growth: a panel data study. *Journal of Economic Structures*, 2023, 12.1: 13.

GUIMARÃES, Inácio Andruski; CHAVES NETO, Anselmo. **Reconhecimento de padrões: metodologias estatísticas em crédito ao consumidor**. *RAE eletrônica*, v. 1, p. 1-14, 2002.

GURGEL, Yara Maria Pereira. **Direitos humanos, princípio da igualdade e não discriminação: sua aplicação às relações de trabalho**. 2007.

LEE, Nicol T.; RESNICK, Paul; BARTON, Genie. **Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms**. Brookings Institution, 2019. Disponível em: <https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>. Acesso em: 02 ago. 2024.

LIMA, Fabiano Guasti, et al. **Aplicação de redes neurais na análise e na concessão de crédito ao consumidor**. *Revista de Administração-RAUSP*, 2009, 44.1: 34-45.

LUMANGO, José Francisco. **Fintechs, regulação e supervisão**. 2020. Tese (Doutorado) – Universidade Nova de Lisboa, Lisboa, 2020.

MANDELA, Nelson. **Struggle is my life**. [S.l.]: Pathfinder Press, 1986.

MITCHELL, Tom M. **Machine learning**. McGraw-Hill Education, 1997. Disponível em: <https://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html>. Acesso em: 02 ago. 2024.

NAÇÕES UNIDAS BRASIL. **Sustainable development goal 3: saúde e bem-estar**. Disponível em: <https://brasil.un.org/pt-br/sdgs/3>. Acesso em: 02 ago. 2024.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Declaração universal dos direitos humanos**. 1948. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 02 ago. 2024.

RIBEIRO, Ricardo S. **Inteligência artificial, direito e equidade algorítmica: discriminações sociais em modelos de machine learning para a tomada de decisão**. Senado Federal, 2022. Disponível em: https://www12.senado.leg.br/ril/edicoes/59/236/ril_v59_n236_p29. Acesso em: 02 ago. 2024.

ROMA, Júlio César. **Os objetivos de desenvolvimento do milênio e sua transição para os objetivos de desenvolvimento sustentável**. *Ciência e Cultura*, v. 71, n. 1, p. 33-39, 2019.

ROSSETTI, Regina; ANGELUCI, Alan. **Ética algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação**. SciELO, 2021. Disponível em: <https://www.scielo.br/j/gal/a/R9F45HyqFZMpQp9BGTfZnyr/?lang=pt>. Acesso em: 02 ago. 2024.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial intelligence: a modern approach**. Pearson, 2016.

SANTOS, Paloma dos. **Transformação digital no sistema bancário: o impacto dos bancos digitais no mercado financeiro no Brasil**. 2023.

TRIVELATO, Bianca Freire; MENDES, Daniel Pacheco; DIAS, Marco Antônio. **A importância do gerenciamento de riscos nas organizações contemporâneas**. *Refas: Revista Fatec Zona Sul*, v. 4, n. 2, p. 3, 2018.

TURING, Alan M. **Computing machinery and intelligence**. UMBC, 1950. Disponível em: <https://courses.cs.umbc.edu/471/papers/turing.pdf>. Acesso em: 02 ago. 2024.

UEDA, Marcos. **Desafios regulatórios no setor financeiro: uma análise das transformações promovidas por fintechs no mercado de meios de pagamento no Brasil**. *Revista de Defesa da Concorrência*, v. 11, n. 2, p. 143-171, 2023.

UNESCO. **Recomendação sobre a ética da inteligência artificial**. UNESCO Digital Library, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por. Acesso em: 02 ago. 2024.

VALLE, Vanice Regina Lírio do; SANTOS, Marcelo Pereira dos. **Governança e compliance na administração direta: ampliando as fronteiras do controle democrático**. *A&C - Revista de Direito Administrativo & Constitucional*, Belo Horizonte, v. 19, n. 75, p. 161–177, 2019. DOI: 10.21056/aec.v20i75.993. Disponível em: <https://www.revistaaec.com/index.php/revistaaec/article/view/993>. Acesso em: 29 set. 2024.

VARGAS, Isadora F.; SANTOS, Nicolas P. **A inclusão financeira por meio de plataformas digitais: experiências de países emergentes e análise do estado da arte no Brasil**. *Revista da Procuradoria-Geral do Banco Central*, 2021. Disponível em: <https://revistapgbc.bcb.gov.br/revista/article/view/1078>. Acesso em: 02 ago. 2024.

VERMA, Mahesh. **An overview of supervised, unsupervised, and reinforcement learning in machine learning**. C# Corner, 2023. Disponível em: <https://www.c-sharpcorner.com/article/an-overview-of-supervised-unsupervised-and-reinforcement-learning-in-machine-ll/>. Acesso em: 02 ago. 2024.

YAN, Jimmy L. **Predição de preços de mercado financeiro baseada na análise de sentimentos e aprendizado profundo**. Repositório Institucional da UFSC, 2023. Disponível em: <https://repositorio.ufsc.br/handle/123456789/253090>. Acesso em: 02 ago. 2024.

YOSHINAGA, Claudia E.; CASTRO, F. Henrique. **Inteligência artificial: a vanguarda das finanças**. *Revista GV Executivo*, 2023. Disponível em: <https://periodicos.fgv.br/gvexecutivo/article/download/89911/84369>. Acesso em: 02 ago. 2024.

GLOSSÁRIO

Definições de acordo com ChatGPT (2024):

Algoritmo de Equidade	Conjunto de técnicas e métodos projetados para garantir que os sistemas de IA sejam justos e não discriminam grupos específicos de indivíduos. Esses algoritmos tentam mitigar o viés nos resultados, promovendo a igualdade de oportunidades.
Análise de Dados	Processo de examinar, limpar e modelar dados com o objetivo de descobrir informações úteis, apoiar a tomada de decisões e identificar padrões.
Aprendizado de Máquina (Machine Learning - ML)	Subcampo da Inteligência Artificial que se concentra no desenvolvimento de algoritmos e modelos que permitem que os sistemas aprendam e façam previsões ou decisões com base em dados. Inclui técnicas como redes neurais, árvores de decisão e aprendizado supervisionado e não supervisionado.
Auditoria de Algoritmo	Processo de revisão e avaliação dos algoritmos e sistemas de IA para identificar e corrigir possíveis vieses, erros ou práticas não conformes.
Big Data	Conjunto de dados extremamente grandes e complexos que não podem ser facilmente gerenciados ou analisados com métodos tradicionais. Geralmente requer ferramentas e técnicas avançadas para processamento e análise.
Compliance	Conformidade com leis, regulamentações, e políticas estabelecidas, especialmente em relação à ética e práticas de negócios.
Crédito Social	Sistema de pontuação que avalia o comportamento financeiro e a confiabilidade dos indivíduos com base

		em seu histórico de crédito e outros dados financeiros.
Desigualdade Digital		Diferença no acesso e uso de tecnologias digitais e internet entre diferentes grupos sociais, econômicos ou geográficos.
Desigualdade de Oportunidades	de	Situação em que certos grupos de indivíduos têm menos acesso a oportunidades e recursos devido a preconceitos ou discriminações sistemáticas.
Direitos Humanos		Conjunto de princípios e normas universais que garantem a dignidade, a liberdade e a igualdade para todos os indivíduos, independentemente de sua origem, raça, gênero, religião ou qualquer outra característica.
Equidade		Princípio que busca garantir que todas as pessoas recebam tratamento justo e igualitário, levando em conta suas necessidades e circunstâncias específicas, para alcançar um resultado equilibrado.
Ética na IA		Conjunto de princípios e diretrizes que visam garantir que o desenvolvimento e uso de tecnologias de inteligência artificial sejam justos, responsáveis e respeitem os direitos humanos.
Exclusão Financeira		Situação em que indivíduos ou grupos são impedidos de acessar serviços financeiros básicos devido a barreiras econômicas, sociais ou tecnológicas.
Governança de Dados		Conjunto de práticas e políticas que garantem que os dados sejam gerenciados e usados de forma eficaz, segura e conforme as normas legais e éticas.
Inclusão Financeira		Acesso igualitário e justo a serviços financeiros e oportunidades econômicas para todos os indivíduos,

independentemente de sua origem, status econômico ou localização geográfica.

Inteligência Artificial (IA)	Área da ciência da computação que se dedica à criação de sistemas e máquinas capazes de realizar tarefas que normalmente requerem inteligência humana, como reconhecimento de padrões, processamento de linguagem natural e tomada de decisão.
Mitigação de Viés	Conjunto de estratégias e técnicas usadas para reduzir ou eliminar o viés em sistemas de IA e algoritmos, garantindo que as decisões tomadas sejam justas e imparciais.
Modelagem Preditiva	Técnica de análise de dados que usa algoritmos e modelos estatísticos para prever futuros eventos ou comportamentos com base em dados históricos.
Políticas Públicas	Diretrizes e ações estabelecidas por governos ou instituições para direcionar e regulamentar áreas específicas da sociedade, visando a promoção do bem-estar público e a resolução de problemas sociais e econômicos.
Regulação	Conjunto de leis, normas e diretrizes que governam a operação de sistemas e práticas em um setor específico, com o objetivo de garantir conformidade, proteção ao consumidor e justiça.
Sistema de Recomendação	Algoritmo de IA utilizado para sugerir produtos, serviços ou informações com base nas preferências e comportamentos dos usuários.
Transparência Algorítmica	Princípio que envolve a abertura sobre como os algoritmos funcionam, incluindo a lógica e os dados usados, para que os usuários e <i>stakeholders</i> possam entender e avaliar como as decisões são tomadas.

Transparência Dados	de Princípio que se refere à clareza sobre como os dados são coletados, utilizados e armazenados por sistemas de IA, permitindo que os usuários compreendam e avaliem o impacto das suas informações pessoais.
Viés Algorítmico	Tendência de sistemas de IA a reproduzir ou amplificar preconceitos presentes nos dados utilizados para seu treinamento. Esse viés pode resultar em decisões injustas ou discriminatórias que afetam negativamente grupos de pessoas.
Viés de Avaliação	Tipo de viés que ocorre quando a forma como um modelo de IA é avaliado ou testado é tendenciosa, o que pode levar a uma avaliação incorreta de seu desempenho e justiça.
Viés de Dados	Tipo de viés que ocorre quando os dados usados para treinar um sistema de IA contêm preconceitos ou desigualdades, levando a resultados injustos ou distorcidos.
Viés de Modelo	Tipo de viés introduzido pela forma como um modelo de IA é construído ou treinado, que pode resultar em decisões enviesadas, mesmo se os dados forem imparciais.

