

I ENCONTRO INTERNACIONAL EM DIREITO E INOVAÇÃO

GOVERNANÇA DIGITAL NA SAÚDE

G721

Governança digital na saúde [Recurso eletrônico on-line] organização I Encontro Internacional em Direito e Inovação: Universidade Católica de Pernambuco – Recife/PE;

Coordenadores: Carol de Oliveira Abud, Vinicius de Negreiros Calado e Marcelo Lamy – Recife: Universidade Católica de Pernambuco, 2025.

Inclui bibliografia

ISBN: 978-65-5274-426-5

Modo de acesso: www.conpedi.org.br em publicações

Tema: Sustentabilidade, Consensualidade, Governança Digital e Inteligência Artificial.

1. Consensualidade. 2. Governança Digital. 3. Inteligência Artificial. I. Encontro Internacional em Direito e Inovação (1:2025 : Recife, PE).

CDU: 34

I ENCONTRO INTERNACIONAL EM DIREITO E INOVAÇÃO

GOVERNANÇA DIGITAL NA SAÚDE

Apresentação

É com imensa satisfação que apresentamos à comunidade acadêmica os frutos colhidos durante o I Encontro Internacional em Direito e Inovação (I EIDI), realizado de 4 a 7 de novembro de 2025.

As páginas que se seguem reúnem as pesquisas que foram aprovadas e apresentadas nos Grupos de Trabalho (GTs) deste evento, que já nasce como um marco para os estudos na intersecção entre o Direito e as novas tecnologias.

O I EIDI nasceu do anseio de criar um fórum qualificado para debater os desafios contemporâneos, sob a égide dos temas da Sustentabilidade, Consensualidade, Governança Digital e Inteligência Artificial.

Promovido pelo Programa de Pós-graduação em Direito e Inovação (PPGDI) da Universidade Católica de Pernambuco (UNICAP), com fomento do Programa de Apoio a Eventos no País (PAEP-20253520241P) e apoio do Conselho Nacional de Pesquisa e Pós-graduação em Direito (CONPEDI), da Escola Judicial do Tribunal de Justiça de Pernambuco - TJPE / ESMape, da Placa Mãe.org, da Escola Superior Dom Helder Câmara (MG), do Mestrado Acadêmico em Direito do CESMAC (AL), do Programa de Pós-Graduação em Direito da Faculdade de Direito de Franca – São Paulo (SP), do Programa de Pós-Graduação em Direito do Centro Universitário 7 de Setembro (UNI7) – Fortaleza (CE), do Programa de Pós-Graduação em Direitos Fundamentais – PPGDF/UNAMA (PA), do Programa de Pós-Graduação em Gestão do Desenvolvimento Local Sustentável (UPE/PE) e do Programa de Pós-Graduação Stricto Sensu da Universidade Santa Cecília (Unisanta) - Mestrado em Direito da Saúde, o evento buscou fomentar um diálogo transdisciplinar e global, conectando pesquisadores em torno de soluções inovadoras para as complexas demandas do século XXI.

A resposta da comunidade acadêmica ao nosso chamado foi, em si, um testemunho da pertinência e da urgência de nossa proposta. Recebemos um volume extraordinário de 148 submissões de resumos expandidos, um número que superou todas as nossas expectativas para uma primeira edição. Após um rigoroso processo de avaliação por pares duplo-cego, que garantiu a impessoalidade e a excelência do processo, 130 trabalhos foram aprovados para apresentação em nossos onze Grupos de Trabalho.

A abrangência do encontro manifestou-se na diversidade de origens, com a valiosa participação de pesquisadores de todas as regiões do Brasil e de colegas da Argentina, Colômbia e Itália, consolidando o caráter internacional de nossos debates.

Este e-book representa, portanto, um recorte desta rica produção intelectual. Cada resumo expandido aqui presente reflete o engajamento, a profundidade e o espírito inovador que permearam as discussões em seu respectivo Grupo de Trabalho. São contribuições que não apenas diagnosticam os desafios atuais, mas que, sobretudo, apontam para novos caminhos, novas interpretações e novas soluções.

Nossa gratidão estende-se a todos que tornaram este projeto possível: aos autores, pela excelência de suas pesquisas; aos coordenadores dos Grupos de Trabalho, pela maestria na condução dos debates; ao Comitê Científico e aos pareceristas, pelo trabalho criterioso e dedicado de avaliação imparcial; e, de forma especial, ao Conselho Nacional de Pesquisa e Pós-graduação em Direito (CONPEDI), cuja parceria foi fundamental para viabilizar este legado editorial.

Que a leitura destas páginas inspire novas reflexões, fomente novas pesquisas e contribua para o avanço contínuo do Direito em uma sociedade em constante transformação. Este é apenas o marco inicial de uma jornada que esperamos continuar a trilhar junto a todos vocês.

Livia Dias Barros

Coordenadora da Comissão Organizadora

Coordenadora do PPGDI/UNICAP

Vinicius de Negreiros Calado

Coordenador do Comitê Científico

Professor Fundador do PPGDI/UNICAP

O PARADOXO DA EFICIÊNCIA: COMO A INTELIGÊNCIA ARTIFICIAL CODIFICA E AMPLIFICA A DISCRIMINAÇÃO ESTRUTURAL

THE PARADOX OF EFFICIENCY: HOW ARTIFICIAL INTELLIGENCE ENCODES AND AMPLIFIES STRUCTURAL DISCRIMINATION

Andressa Felix Lisboa ¹
Carolina Cruz Rodriguez Coelho ²
Carol de Oliveira Abud ³

Resumo

A Inteligência Artificial (IA) emerge como uma força transformadora, prometendo otimizar processos e eliminar a subjetividade humana na tomada de decisão. Contudo, esse potencial esbarra em um paradoxo central: em vez de erradicar o preconceito, os sistemas algorítmicos tendem a codificar e amplificar discriminações estruturais. O artigo analisa os mecanismos pelos quais o viés se infiltra nos sistemas de IA, desde os dados de treinamento enviesados até o uso de variáveis proxy, e examina as limitações regulatórias do AI Act da União Europeia. Adota-se pesquisa qualitativa, de caráter descritivo e analítico, baseada em revisão bibliográfica e análise normativa. Os resultados indicam que a discriminação algorítmica decorre de processos técnicos e sociais que reproduzem desigualdades históricas, evidenciando a insuficiência da regulação baseada em risco. Conclui-se que o enfrentamento efetivo desse problema exige uma abordagem orientada pela justiça algorítmica, centrada na equidade e nos direitos fundamentais.

Palavras-chave: Discriminação algorítmica, Direitos fundamentais, Justiça algorítmica, Regulação tecnológica

Abstract/Resumen/Résumé

Artificial Intelligence (AI) emerges as a transformative force, promising to optimize processes and eliminate human subjectivity in decision-making. However, this potential encounters a central paradox: instead of eradicating bias, algorithmic systems tend to encode and amplify structural discriminations. The article analyzes the mechanisms through which bias infiltrates AI systems, from biased training data to the use of proxy variables, and examines the regulatory limitations of the European Union's AI Act. A qualitative research

¹ Advogada e professora, especialista em Seguridade Social. Mestre em Direito da Saúde pela UNISANTA (bolsista CAPES) e aluna especial do Doutorado em Direito (UFBA). Atua nas áreas previdenciária e trabalhista.

² Advogada, bacharel em Direito pela UNISANTA (2022). Pós-graduanda em Direito Notarial e Registral. Pesquisadora no projeto Marco Regulatório dos Medicamentos e membro do grupo CNPq/UNISANTA.

³ Professora orientadora; Doutoranda em Ciência e Tecnologia Ambiental e Mestra em Direito da Saúde pela Universidade Santa Cecília; Advogada.

approach is adopted, of descriptive and analytical nature, based on bibliographic review and normative analysis. The results indicate that algorithmic discrimination arises from technical and social processes that reproduce historical inequalities, highlighting the insufficiency of risk-based regulation. It is concluded that effectively addressing this problem requires an approach guided by algorithmic justice, centered on equity and fundamental rights.

Keywords/Palabras-claves/Mots-clés: Algorithmic discrimination, Fundamental rights, Algorithmic justice, Technological regulation

INTRODUÇÃO

Assim como a Revolução Industrial redefiniu a produção e o trabalho, a revolução da Inteligência Artificial (IA) promete reconfigurar os alicerces da sociedade contemporânea. A capacidade de processar volumes massivos de dados (Big Data) e identificar padrões complexos impulsionou a adoção de sistemas automatizados em esferas de decisão que afetam diretamente a vida humana: seja na triagem de currículos para emprego, na concessão de crédito, na administração da justiça ou no diagnóstico médico.

A promessa subjacente a essa transição é a da eficiência e, sobretudo, da objetividade. A expectativa é que um algoritmo, livre das emoções e preconceitos estruturais humanos, possa tomar decisões mais justas e precisas. No entanto, sob a superfície dessa eficiência, cresce a evidência de que a IA está se tornando um vetor poderoso para a replicação de desigualdades sistêmicas.

O problema fundamental é que a IA não aprende no vácuo; ela aprende a partir do mundo como ele é (ou foi). Os algoritmos são treinados com dados históricos que, por sua vez, estão repletos de preconceitos sociais, raciais, de gênero e econômicos acumulados ao longo de décadas. Um sistema de IA treinado para prever o “sucesso profissional” com base em dados de contratação dos últimos 50 anos aprenderá, inevitavelmente, que “sucesso” está correlacionado a perfis historicamente dominantes, perpetuando a exclusão de grupos minorizados.

A discriminação algorítmica, portanto, não é frequentemente um erro intencional ou um “bug” no sistema; é o sistema funcionando exatamente como projetado, refletindo com precisão os vieses presentes nos dados. Isso levanta questões éticas e jurídicas cruciais: como podemos garantir os princípios da equidade e da não discriminação quando as ferramentas de decisão são inerentemente enviesadas?

Diante desse cenário, este artigo busca compreender a anatomia da discriminação algorítmica. Em vez de tratar o viés como um problema monolítico, exploraremos seus diferentes mecanismos, desde os dados de treinamento até o design do modelo, e analisaremos criticamente as limitações das abordagens regulatórias atuais para mitigar esses riscos profundos e estruturais.

O objetivo geral deste estudo é analisar de forma crítica os mecanismos pelos quais o viés discriminatório se infiltra nos sistemas de inteligência artificial, desde a formação de dados enviesados até o uso de variáveis proxy que produzem efeitos discriminatórios indiretos, examinando as limitações das regulações emergentes, especialmente do AI Act da União Europeia.

Como desdobramentos específicos, o artigo busca: (i) identificar as formas e origens do viés algorítmico, relacionando-as a processos sociais e institucionais de discriminação estrutural; (ii) avaliar como a utilização de dados históricos e proxies contribui para a reprodução e amplificação de desigualdades nos sistemas de IA; (iii) investigar os fundamentos éticos e jurídicos das principais iniciativas regulatórias, com destaque para o modelo europeu; (iv) discutir a insuficiência das abordagens baseadas em risco e propor a transição conceitual da “IA de confiança” para uma “IA justa”, comprometida com a equidade e os direitos fundamentais.

A pesquisa é de natureza qualitativa, com caráter descritivo-analítico e exploratório, desenvolvida por meio de revisão bibliográfica e análise normativa do AI Act. Foram levantadas fontes primárias, como o regulamento e as diretrizes europeias, e fontes secundárias, representadas por autores como Moreira, Moraes e Lisboa. A seleção bibliográfica considerou publicações compreendidas entre os anos de 2019 e 2025, provenientes de bases científicas reconhecidas. A abordagem teórica adotada é interdisciplinar (Lamy, 2020), integrando os campos do Direito, da Ética e da Ciência de Dados. Como limitação do estudo, destaca-se a ausência de análise empírica de casos brasileiros e o foco concentrado no contexto europeu.

RESULTADOS E DISCUSSÃO

Para regular efetivamente a IA, é preciso primeiro compreender como a discriminação se manifesta. Ela raramente é explícita. Em vez disso, opera através de mecanismos técnicos e estatísticos complexos que obscurecem sua origem e impacto.

Observa-se que, a discriminação é frequentemente associada a concepções tradicionais que a definem como um ato intencional e arbitrário. Moreira, apresenta a definição clássica de discriminação direta como sendo: “[...] uma ação intencional e arbitrária baseada em um critério de tratamento ilegítimo, o que pode colocar as pessoas em uma situação de desvantagem temporária ou duradoura” (Moreira, 2019, p. 29). Essa visão, no entanto, é apresentada como uma compreensão limitada do fenômeno, que não abarca suas manifestações mais complexas e estruturais.

Moreira expande essa definição ao argumentar que a discriminação possui uma “natureza dinâmica e múltipla” (Moreira, 2019, p. 24). Ela não se restringe a atos “abertamente discriminatórios”, mas se manifesta de formas sutis, como as microagressões, que incluem microinsultos definidos como “formas de comunicação que demonstram de maneira expressa ou encoberta uma ausência de sensibilidade à experiência, à tradição ou à identidade cultural” (Moreira, 2019, p. 37), podendo ser “não propositais”. Além da ação individual, a discriminação

assume uma dimensão institucional, que “designa práticas institucionais que podem ou não levar necessariamente a raça em consideração, mas que mesmo assim afetam certos grupos raciais de forma negativa” (Moreira, 2019, p. 34-35).

Ao que tange o viés algorítmico, são os dados de treinamento. Se um sistema é alimentado com dados que refletem um passado discriminatório, ele inevitavelmente aprenderá e replicará essas tendências. Por exemplo, sistemas de policiamento preditivo, quando treinados com dados históricos de prisões, tendem a direcionar mais policiamento para bairros de minorias, não porque haja mais crimes, mas porque historicamente houve mais vigilância e prisões nessas áreas.

Para Moraes e Lisboa, o treinamento de sistemas com “dados enviesados podem perpetuar e até exacerbar discriminações existentes” (Moraes; Lisboa, 2024), criando “impactos negativos substanciais para indivíduos e grupos vulneráveis” (Moraes; Lisboa, 2024).

Sob a perspectiva de Moreira, observa-se que o que compõe esses “dados enviesados”: são o registro histórico de “estereótipos” e “estigmas” que, por sua vez, são “meios de legitimação de arranjos sociais excludentes” (Moreira, 2019, p.42).

Desta forma, quando um algoritmo é treinado com dados históricos, ele não está aprendendo uma realidade neutra, mas sim internalizando o “racismo simbólico”, definido por Moreira como “construções culturais que estruturam a forma como minorias raciais são representadas” (Moreira, 2019, p. 34).

A IA se torna um veículo moderno para a perpetuação de vieses justamente porque opera de forma similar ao “racismo institucional” e às “microagressões” descritas por Moreira (2019, p. 37).

O racismo institucional, segundo o autor, pode ocorrer “sem um caráter necessariamente intencional”, afetando grupos de forma negativa através do “funcionamento normal de instituições”. Um algoritmo que reproduz o viés opera exatamente dessa maneira: ele não tem “intenção”, mas seu funcionamento normal replica a exclusão (Moreira, 2019, p. 34-35).

Isso se alinha também ao conceito de “microinsultos”, que podem ser “não propositais”, mas “demonstram de maneira [...] encoberta uma ausência de sensibilidade” (Moreira, 2019, p. 37). Moreira aponta que estereótipos possuem uma “dimensão prescritiva”, que designa “o lugar que os indivíduos que pertencem a certos segmentos podem ocupar” (Moreira, 2019, p. 43).

Moraes e Lisboa, apontam a problemática com sistemas de “risco elevado” usados em “recrutamento” ou “concessão de empréstimos”, pois são precisamente esses sistemas que, ao

automatizar o estereótipo prescritivo, definem o lugar social e as oportunidades materiais dos indivíduos (Moraes; Lisboa, 2024).

Isso cria um ciclo de retroalimentação perigoso, onde, por exemplo: a IA prevê o crime com base em dados enviesados, a polícia é enviada para aquela área, mais prisões (mesmo que por infrações menores) ocorrem, e esses novos dados “confirmam” a previsão original do algoritmo. O viés não é apenas replicado; ele é amplificado e justificado sob um verniz de objetividade tecnológica.

Na medicina, esse risco é igualmente presente. Como apontado no debate sobre a regulação europeia, um algoritmo de diagnóstico de câncer de pele treinado predominantemente com imagens de pele clara pode falhar em identificar melanomas em pacientes de pele escura, exacerbando desigualdades na saúde.

Um dos mecanismos mais insidiosos de discriminação algorítmica é o uso de proxies (variáveis substitutas). Os legisladores frequentemente proíbem o uso de “variáveis sensíveis”, como raça, gênero ou religião, na tomada de decisão. No entanto, um algoritmo de *deep learning* pode facilmente identificar correlações entre dados aparentemente neutros e essas categorias protegidas.

O uso de variáveis *proxy* em sistemas de inteligência artificial, ao reproduzir desigualdades históricas sob a aparência de neutralidade técnica, viola os princípios de dignidade humana, igualdade e não discriminação consagrados na Carta dos Direitos Fundamentais da União Europeia, que estabelece como imperativo jurídico o respeito incondicional aos direitos e liberdades fundamentais em todas as aplicações tecnológicas (União Europeia, 2000).

Por exemplo, um algoritmo de concessão de crédito pode não usar “raça” como variável, mas pode usar o CEP (código postal), o histórico de compras, o tipo de escola frequentada ou até mesmo os padrões de digitação. Se esses proxies estiverem estatisticamente correlacionados com a raça ou a classe social do indivíduo, o algoritmo pode efetivamente discriminar sem nunca “saber” a raça da pessoa.

Isso torna a auditoria de viés extremamente difícil. O desenvolvedor do sistema pode alegar que o modelo é neutro, pois usa apenas dados financeiros ou comportamentais, quando, na prática, está engajado em uma forma sofisticada de discriminação indireta.

Muitos modelos de IA mais avançados, especialmente redes neurais complexas, operam como “caixas-pretas”. Isso significa que, embora possam produzir resultados altamente precisos, seus processos internos de decisão são impenetráveis ou incompreensíveis até mesmo para seus criadores.

Essa falta de “explicabilidade” colide diretamente com os princípios jurídicos fundamentais, como o devido processo legal e o direito à contestação. Se um indivíduo tem um pedido de empréstimo ou um benefício social negado por uma IA, como ele pode contestar essa decisão se ninguém consegue explicar por que ela foi tomada?

A proteção dos direitos e liberdades fundamentais é o objetivo da abordagem europeia à IA tendo como alicerce a Carta dos Direitos Fundamentais da União Europeia para garantir que a tecnologia não infrinja os direitos humanos (Sousa, 2020). Além dessa base jurídica, o regulamento observou diretrizes anteriores publicadas pela Comissão Europeia em abril de 2019, conhecidas como “Orientações Éticas para uma IA de Confiança” (2019, p. 48-50).

O referido documento trabalha o conceito de “IA de confiança” em quatro princípios éticos e sete requisitos essenciais. Os princípios éticos chave são (1) o respeito pela autonomia humana, que postula a manutenção do controle humano sobre os sistemas de IA para proteger os indivíduos de manipulação; (2) a prevenção de danos, focada na segurança e bem-estar da sociedade; (3) a equidade, que exige o desenvolvimento e operação justos e imparciais para prevenir discriminação; e (4) a explicabilidade, definida como a capacidade de compreender e justificar as decisões da IA, assegurando transparência e o direito de contestação (Orientações Éticas para uma IA de Confiança, 2019, p. 48-50).

No entanto, a complexidade técnica desses sistemas pode tornar a supervisão humana superficial, onde o humano simplesmente “carimba” a decisão da máquina sem uma compreensão real de sua lógica interna, falhando em identificar o viés oculto.

CONSIDERAÇÕES FINAIS

Os resultados indicam que a discriminação algorítmica não decorre de falhas isoladas, mas da própria lógica de funcionamento dos sistemas de inteligência artificial, que aprendem e reproduzem padrões históricos de desigualdade. Verificou-se que dados de treinamento enviesados e o uso de variáveis substitutas, como localização ou histórico de consumo, permitem que algoritmos discriminem mesmo sem recorrer diretamente a critérios sensíveis como raça ou gênero. Essa dinâmica cria ciclos de retroalimentação que reforçam a exclusão de grupos minorizados e dificultam a detecção e a correção do viés. Observou-se também que a complexidade técnica dos modelos e a opacidade de suas decisões comprometem a transparência e o direito de contestação, limitando a efetividade dos princípios jurídicos de equidade e devido processo. Constatou-se, por fim, que as regulações baseadas apenas na classificação de risco, como o AI Act europeu, são insuficientes para enfrentar formas

estruturais e sutis de discriminação, evidenciando a necessidade de um paradigma de justiça algorítmica que una controle técnico, ética e direitos fundamentais.

A pesquisa permitiu constatar que os sistemas de inteligência artificial, ao processarem dados históricos marcados por desigualdades estruturais, reproduzem e amplificam discriminações de forma automatizada. Os resultados parciais demonstram que a discriminação algorítmica não se limita a falhas técnicas, mas emerge de escolhas de design e treinamento que refletem contextos sociais hierarquizados. A análise também revelou que, embora o AI Act da União Europeia represente um avanço normativo, sua abordagem de risco é limitada para enfrentar mecanismos sutis de exclusão, como o uso de proxies e a falta de explicabilidade algorítmica. Esses achados reforçam a urgência de repensar a governança tecnológica sob o paradigma de uma inteligência artificial justa, orientada pelos direitos humanos e pela equidade.

O avanço da inteligência artificial trouxe consigo oportunidades e desafios, sobretudo quanto à sua aplicação em processos de classificação de pessoas e grupos. À medida que esses sistemas se integram às rotinas sociais e institucionais, torna-se indispensável que sua implementação se fundamente em princípios éticos e jurídicos robustos. A não discriminação, a transparência e a explicabilidade configuram pilares indispensáveis para garantir o uso justo e responsável da IA exigindo mecanismos que assegurem o direito de compreensão e contestação das decisões automatizadas. Contudo, a regulação isolada não basta: é essencial que desenvolvedores, pesquisadores e usuários cultivem uma cultura de responsabilidade em todas as fases do ciclo de vida tecnológico, adotando práticas de design inclusivas, avaliações de impacto ético e mecanismos efetivos de supervisão.

REFERÊNCIAS

- COMISSÃO EUROPEIA. Direção-Geral das Redes de Comunicação, Conteúdos e Tecnologias. **Orientações éticas para uma IA de confiança**. [S.l.]: Serviço das Publicações, 2019. Disponível em: <https://data.europa.eu/doi/10.2759/2686>. Acesso em 27 de out. 2025
- LAMY, Marcelo. **Metodologia da pesquisa**. Técnicas de investigação, argumentação e redação. 2ª ed. revista, atualizada e ampliada. São Paulo, SP: Matrioska Editora, 2020.
- MORAES, Alexandre Rocha Almeida de; LISBOA, Andressa Felix. Regulamentação da Inteligência Artificial na União Europeia: estrutura ética, classificação de riscos e possíveis reflexos na medicina. **Unisanta Law and Social Science**, Santos, v. 13, n. 2, p. 16-29, 2024. DOI: 10.5281/zenodo.14174188. Disponível em: <https://periodicos.unisanta.br/LSS/article/view/2473>. Acesso em 27 de out. 2025.
- MOREIRA, Adilson. **Racismo recreativo**. São Paulo: Sueli Carneiro; Pólen, 2019. (Feminismos Plurais / coordenação de Djamila Ribeiro)
- SOUSA, Maria Eliane Alves. Direitos humanos e princípios comuns entre inteligência artificial e direito à saúde. **Cadernos Ibero-Americanos de Direito Sanitário**, [S. l.], v. 9, n.

3, p. 26–48, 2020. DOI: 10.17566/ciads.v9i3.658. Disponível em: <https://www.cadernos.prodisa.fiocruz.br/index.php/cadernos/article/view/658> . Acesso em: 24 abr. 2024.

UNIÃO EUROPEIA. **Carta dos Direitos Fundamentais da União Europeia**. Jornal Oficial das Comunidades Europeias, C 364/1, 18 de dezembro de 2000. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:12012P/TXT>. Acesso em: 27 out. 2025.