Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

Artigo de pesquisa

Marjori Klinczak
Unifatec
ORCiD 0009-0009-2028-8359

José Simão de Paula Pinto
Universidade Federal do Paraná
ORCiD 0000-0002-5023-437X

GOVERNANÇA ALGORÍTMICA E ÉTICA NA INTELIGÊNCIA: UMA ANÁLISE DAS INICIATIVAS EMERGENTES

https://doi.org/10.58960/rbi.2026.21.291

Klinczak, Marjori e José Simão de Paula Pinto. 2026. "Governança Algorítmica e Ética na Inteligência: uma análise das iniciativas emergentes". *Revista Brasileira de Inteligência* (ABIN) 21: e2026.21.291. https://doi.org/10.58960/rbi.2026.21.291.

Recebido em 13/10/2025
Aprovado em 14/05/2026
Publicado em 13/07/2026

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

GOVERNANÇA ALGORÍTMICA E ÉTICA NA INTELIGÊNCIA: UMA ANÁLISE DAS INICIATIVAS EMERGENTES

RESUMO

Os sistemas de informação e IA (inteligência artificial) estão presentes em diversos aspectos da vida cotidiana, como comunicação, trabalho e estudo. Porém, essa exposição digital possibilita a coleta massiva de dados sobre hábitos e escolhas pessoais, muitas vezes sem consentimento explícito, que podem ser usados para criação de perfis, sistemas de recomendação e prevenção de crimes. No entanto, os sistemas que processam esses dados não estão imunes a vieses, podendo afetar direitos humanos, privacidade e liberdade de expressão. Nesse contexto, a governança algorítmica busca promover maior transparência, ética e responsabilização nas decisões automatizadas. Tem-se então como objetivo analisar as iniciativas de governança algorítmica e de IA voltadas à mitigação de problemas decorrentes de vieses e discriminação, através de uma metodologia bibliográfica e exploratória.

Palavras-chave: *governança algorítmica, inteligência artificial, tomada de decisão, algoritmos.*

ALGORITHMIC GOVERNANCE AND ETHICS IN INTELLIGENCE: AN ANALYSIS OF EMERGING INITIATIVES

ABSTRACT

Information systems and artificial intelligence are present in various aspects of daily life, such as communication, work, and education. However, this digital exposure enables the massive collection of data on personal habits and choices, often without explicit consent, which can be used for the creation of profiles, recommendation systems, and crime prevention. Nevertheless, the systems that process these data are not immune to biases, potentially affecting human rights, privacy, and freedom of expression. In this context, algorithmic governance seeks to promote greater transparency, ethics, and accountability in automated decision-making. The objective of this study is to analyze algorithmic and AI governance initiatives aimed at mitigating problems arising from biases and discrimination, through a bibliographic and exploratory methodology.

Keywords: *algorithmic governance, artificial intelligence, decision-making, algorithms.*

GOBERNANZA ALGORÍTMICA Y ÉTICA EN LA INTELIGENCIA: UN ANÁLISIS DE LAS INICIATIVAS EMERGENTES

RESUMEN

Los sistemas de información y la inteligencia artificial (IA) están presentes en diversos aspectos de la vida cotidiana, como la comunicación, el trabajo y el estudio. Sin embargo, esta exposición digital posibilita la recopilación masiva de datos sobre los hábitos y las decisiones personales, muchas veces sin el consentimiento explícito de los usuarios, los cuales pueden utilizarse para la elaboración de perfiles, los sistemas de recomendación y la prevención del delito. No obstante, los sistemas que procesan estos datos no son inmunes a los sesgos, lo que puede afectar los derechos humanos, la privacidad y la libertad de expresión. En este contexto, la gobernanza algorítmica busca promover una mayor transparencia, ética y rendición de cuentas en las decisiones automatizadas. En este sentido, el objetivo de este trabajo es analizar las iniciativas de gobernanza algorítmica y de inteligencia artificial orientadas a la mitigación de los problemas derivados de los sesgos y la discriminación, mediante una metodología bibliográfica y exploratoria.

Palabras clave: *gobernanza algorítmica; inteligencia artificial; toma de decisiones; algoritmos.*

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

Introdução

É inegável as vantagens e possibilidades que a inteligência artificial têm proporcionado atualmente, conforme Lovatto (2024), mudando inclusive como nos comunicamos e resolvemos diversas tarefas do nosso dia a dia, como reuniões a distância, ensino online, pagamentos sem enfrentar filas e compras de produtos variados, desde alimentos até vestuário.

Além disso, esses sistemas nos permitem fazer escolhas com base na pontuação calculada por sistemas, predizer o comportamento das pessoas, suas preferências e opiniões, o que tem sido utilizado no combate a fraudes, evasão de impostos, predição de doenças, direção de carros autônomos, controle de fronteiras, prevenção ao terrorismo e à criminalidade, bem como na recomendação de produtos, entre outros, conforme Katzenbach e Ulbricht (2019).

Contudo, a eficácia desses resultados está diretamente ligada a qualidade dos dados utilizados para treinamento dos algoritmos, e falhas nesse ponto podem trazer consequências e distorções graves, como as enfrentadas por grandes empresas como Google e Microsoft, por exemplo. A Google em 2016 marcou pessoas de pele negra como gorilas, os carros autônomos têm, em geral, uma maior chance de atropelar pessoas negras e também a Tay, uma inteligência artificial criada pela Microsoft, precisou ser retirada do ar 24h depois de lançada no antigo Twitter, por estar proferindo discursos de ódio, racistas e sexistas (Junqueira e Botelho-Francisco 2021; Issar e Aneesh 2021).

Pouco também se fala sobre o risco associados à criação de perfis de usuários que podem gerar discriminação, dados com vieses, manipulação do consumo e redução de oportunidades entre grupos vulneráveis, de acordo com Lyon (2014) e Issar e Aneesh (2021), infringindo direitos humanos e a liberdade de expressão. Esses desafios colocam em pauta não apenas o campo tecnológico, mas também a própria atividade de inteligência, que precisa considerar como algoritmos podem influenciar a coleta, a análise e a interpretação de informações, afetando processos de decisão e segurança nacional, como por exemplo, no caso das eleições.

Esse crescimento exponencial dos dados disponíveis exige mecanismos capazes de garantir que sua utilização ocorra de forma ética e transparente, especialmente quando associada à inteligência artificial (Errey-Hammond 2024), pois essas novas tecnologias têm transformado a forma como informações são coletadas, armazenadas, analisadas e compartilhadas, ampliando o potencial das atividades de inteligência, mas também os desafios relacio-

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

nados à segurança nacional, à privacidade e à governança dos dados (Boyd e Crawford 2012; Kitchin 2014; Mayer-Schönberger e Ramge 2018). Nesse contexto, os algoritmos e aplicações não se caracterizam mais apenas pelo grande volume de dados que podem processar, mas pela capacidade de integrá-los, correlacioná-los e extrair conhecimento para apoiar a tomada de decisão (Boyd e Crawford 2012), visto que a mesma se torna mais complexa (Stenslie 2022), o que reforça a necessidade de mecanismos de governança algorítmica que assegurem transparência, responsabilização e respeito aos direitos fundamentais.

Diante disso, o objetivo geral deste trabalho é realizar uma análise teórica e empírica das iniciativas voltadas à governança algorítmica, explorando como elas buscam mitigar riscos éticos, legais e sociais advindos do uso de sistemas automatizados. Os objetivos específicos são: (i) definir o conceito de governança algorítmica; (ii) mapear os principais cenários onde ela se insere, incluindo sua relação com a desinformação e a segurança cibernética; (iii) e por fim, examinar, na literatura e na legislação nacional e internacional, iniciativas que buscam enfrentar os desafios éticos e legais impostos pelos algoritmos às práticas de inteligência.

A metodologia adotada é a bibliográfica e exploratória, envolvendo pesquisa teórica em livros e artigos científicos, onde busca-se identificar os pontos críticos em que a governança algorítmica se conecta com os estudos de inteligência, sobretudo no que se refere à proteção da privacidade, ao equilíbrio entre segurança e liberdades civis e às implicações para a soberania e a segurança cibernética. Quanto ao uso de inteligência artificial, fez-se uso do Chat GPT para conversão das referências que estavam no formato ABNT para o formato da revista.

Governança Algorítmica

De acordo com Katzenbach e Ulbricht (2019), o conceito de governança algorítmica tem ganho destaque na última década, além de ter se tornado um novo campo de pesquisa nas ciências sociais (Aneesh 2006). Sua definição é a de uma forma de ordem social que demonstra uma certa coordenação entre os atores, baseada em regras e incorporada em sistemas e procedimentos de informação (Katzenbach e Ulbricht 2019). Apesar disso, o termo foi utilizado pela primeira vez em 2009, por Rouvroy e Berns (2009), e antes disso era mencionado como governança social por Müller-Birn *et al*. (2013).

Seu uso como conhecemos e referindo-se a uma regulamentação dos al-

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

goritmos foi introduzida por Tim O'Reilly (2013), destacando a eficiência dos espaços governados automaticamente e ao mesmo tempo ignorando a despolitização dos mesmos quando se delega as soluções tecnológicas a tomada de decisão. Isso ocorre pois os algoritmos contribuem para uma reorganização da ordem social quando os modelos matemáticos e demais procedimentos sistêmicos apresentam uma nova forma de quantificar e classificar a ordem social, conforme Katzenbach e Ulbricht (2019), sem deixar exatamente claro como os resultados foram gerados ou se provêm de uma base de dados com vieses.

Isso acontece pois os sistemas trabalham com algoritmos que possuem diferentes graus de transparência, e quanto mais automatizado forem, menor o nível de intervenção humana, sendo que muitos algoritmos nem mesmo permitem que se identifique como se chegou a determinado resultado, sendo conhecidos como 'caixa-preta' (*black-box*) (Pasquale 2015). Com isso, a governança algorítmica pode variar dependendo das leis e regulamentações de cada localidade, segundo Issar e Aneesh (2021).

Segundo Doneda e Almeida (2016), os algoritmos podem ser considerados como um conjunto de instruções cujo objetivo é a realização de alguma tarefa, que é produzido a partir de uma entrada, e que gera uma saída correspondente. Eles estão presentes em diversos tipos de sistemas, com diferentes complexidades, principalmente se forem utilizados em conjunto com outras técnicas como aprendizado de máquina ou inteligência artificial, que quando recebem como entrada uma grande quantidade de dados (big data), conseguem fazer diversos relacionamentos e trazer insights (percepções) sobre esses dados, que não seriam facilmente reconhecidos sem o uso desses métodos de alta capacidade relacional, o que, muitas vezes, torna difícil ou impossível refazer o fluxo que o levou a tomar determinada decisão. Quanto a isso, Stenslie (2022) menciona que quanto maior a automação, maior deve ser a governança.

A própria inteligência artificial, que está em grande destaque atualmente devido à criação do Chat GPT pela Open IA e o Gemini pela Google, é uma forma de algoritmo com maior complexidade, e de acordo com Russel e Norvig (1995), uma de suas definições é que ela faz 'tudo certo' com base nos dados que tem, logo se a base de dados utilizada não possuir diversidade e variedade, não for treinada adequadamente, possuir dados não verídicos ou com vieses, o impacto no resultado será direto. Como exemplos tem-se os seguintes casos: a Google que em 2016 catalogou pessoas de pele escura como gorilas, os carros autônomos que têm uma maior chance de atropelar

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

pessoas negras e a Tay, uma inteligência artificial criada pela Microsoft e retirada do ar 24h depois de lançada no antigo Twitter, por ter sido 'ensinada' a proferir discursos de ódio (Junqueira e Botelho-Francisco 2021; Issar e Aneeh 2021).

Assim, faz-se necessário uma solução que faça a coleta de dados de qualidade e depois realize diversos testes, envolvendo várias partes da base de dados, o que é um dos itens que caracteriza um *big data*, ou seja, um grande conjunto de dados com diferentes formas de informação (textos, imagens, logs de sistema), e como muitas vezes essa coleta é feita de forma automática, traz-se para as bases de dados vieses já existentes nos meios físicos, como, por exemplo, aqueles presentes na criação de bases de dados de faces, conforme demonstrado na Figura 1.

Figura 1 – Exemplo de base de dados utilizada para treinamento de algoritmos de reconhecimento facial.

Fonte: Kaggle (2025).

Observa-se na base de dados mencionada um padrão nas imagens, que apesar de ser formada por 13 mil imagens coletadas da Internet, representa na prática pouca diversidade, o que pode gerar um viés no treinamento dos modelos de IA, já que a maior parte das imagens representa pessoas de pele clara e que com isso poder-se-ia atribuir maior relevância a esse perfil em detrimento de outros.

Issar e Aneeh (2021) apresentam os problemas atrelados a governança algorítmica agrupados em 3 categorias: vigilância, enviesamento dos dados e identidade, conforme Figura 2. Seu impacto torna-se óbvio em momentos como eleições, campanhas de saúde pública, entre outros cenários de pos-

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

sível manipulação de opinião pública. Com isso, a transparência na IA e na tomada de decisão permite que os consumidores finais possam entender como e por que determinadas escolhas foram feitas, reduzindo os riscos legais para as empresas, vieses e alucinações algorítmicas, e torna-se possível demonstrar a inexistência de decisões discriminatórias (Lovatto 2024). Ford e Vogel (2025) ainda ressaltam que a supervisão humana permanece indispensável, especialmente em decisões de elevado impacto social, reforçando a necessidade de mecanismos de governança capazes de assegurar transparência e responsabilização.

Figura 2 - Agrupamento dos problemas da governança algorítmica

	Estado	Mercado	Consequências
Problema do poder			
Vigilância	Perda de direitos	Manipulação da assimetria das informações	Domínio (GPS do telefone sempre ligado*)
Problema da discriminação			
Viéses	Bem-estar e justiça (oportunidades iguais)	Reduz a oportunidade no mercado de trabalho ou compra de itens	Redução de oportunidades
Problema de identificação			
Identidade	Criação de perfis (score)	Perfis financeiros, médicos	Perfis nas mídias sociais

Fonte: Adaptado de Issar e Aneesh (2021).

Entre os diversos desafios enfrentados pela governança algorítmica, destaca-se a desinformação, que representa um problema de segurança informacional e um campo sensível para a atividade de inteligência contemporânea., devido a utilização de algoritmos para criação e divulgação de informações falsas de forma massiva, que com o advento da popularização das redes sociais ganha uma escala muito maior. Prova disso é que o Facebook, conforme mencionado por Moreira (2020), atuou como um facilitador na disseminação da desinformação relacionado a saúde em cinco países: Estados Unidos, Reino Unido, França, Alemanha e Itália, fazendo com que as postagens com informações falsas tivessem um alcance de 3,8 bilhões de acessos entre 2019 e 2020.

A desinformação pode então ser caracterizada como uma informação enganosa, inexata ou falsa, sendo um campo crítico para a área de inteligência, criada com intenção de causar prejuízo e envolve um conjunto de ações intencionais para moldar a opinião pública com o interesse de quem a está produzindo, com isso, muitas vezes os fatos são distorcidos ou apresentados fora de contexto, segundo Brisola e Bezerra (2020). E segundo de Mello e Schneider (2021), as pessoas são programadas para responderem a gatilhos emocionais,

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

o que muitas vezes os leva ao compartilhamento dessas informações falsas.

Elas ainda podem ser caracterizadas das diferentes formas, de acordo com Volkoff (2004), misinformation, disinformation e malinformation. A misinformation é a informação falsa, imprecisa ou enganosa, onde não há nenhuma intenção de prejudicar, ao contrário da disinformation, que possui como objetivo causar algum prejuízo. Por fim, a malinformation é a informação falsa, imprecisa ou enganosa ligada aos discursos de ódio, onde existe o objetivo de causar prejuízo a algum grupo específico.

Há ainda o contexto da infodemia, relacionado ao imenso volume de informação falsa referente a um determinado tópico, dificultando a busca pelas informações íntegras, conforme Posetti e Bontcheva (2020). Como exemplos desse cenário podem ser citados a pandemia causada pela Covid-19 e as eleições brasileiras e americanas.

Os autores de Mello e Schneider (2021) reforçam que o pensamento crítico e análise das informações seriam suficientes para barrar a disseminação de notícias falsas, porém, como elas são feitas para sensibilizar o lado emocional do ser humano isso não é tão simples, sem falar nos grupos vulneráveis que nem sempre têm o conhecimento ou instrumentos para conferir se uma notícia é de fato verdadeira ou falsa, ou na infodemia, que dificulta o trabalho de checar a veracidade dos dados.

Katzenbach e Ulbricht (2019) também mencionam que as plataformas digitais têm procurado retirar do ar conteúdos inadequados, porém não é transparente a forma como fazem isso, o tempo que tal ação leva (sem contar processos judiciais exigindo a retirada do conteúdo do ar) e como essa decisão é tomada, principalmente no que se refere a discursos de ódio. O sistema do Youtube tem bloqueado conteúdos que fazem uso de músicas e vídeos com direitos autorais, mas ainda assim faz-se necessário a implementação de sistemas que consigam identificar discursos de ódio e campanhas de desinformação, pois devido ao massivo uso dessas plataformas, a moderação humana é inviável.

Tem-se dessa forma os modelos de governança algorítmica voltadas à IA, baseados em 4 pilares: rastreabilidade (possibilidade de acompanhar todo o fluxo de dados e decisões de um algoritmo, permitindo auditoria e verificação de sua origem e processamento), explicabilidade (habilidade de um sistema de IA de fornecer justificativas compreensíveis sobre como e por que certas decisões foram tomadas), confiabilidade (grau em que um algoritmo produz resultados consistentes, precisos e previsíveis em diferentes condições de

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

uso) e responsabilidade (obrigação de identificar e responsabilizar indivíduos ou organizações pelos impactos e decisões resultantes do uso de algoritmos), e os modelos de governança voltados ao enfrentamento da desinformação, baseados na autorregulação (quando empresas ou plataformas definem e aplicam suas próprias regras e políticas para garantir uso ético e seguro de algoritmos e dados, sendo muito utilizado pelas grandes empresas como Google e Meta), regulação estatal (intervenção de governos para criar leis, normas e mecanismos de fiscalização que orientem o desenvolvimento e uso responsável de algoritmos e IA) e normas internacionais (diretrizes e padrões elaborados por organismos globais ou multilaterais para harmonizar práticas de governança de IA e proteção de direitos em diferentes países) (Lovatto 2024). Porém, elas ainda têm como desafio lidar com a grande quantidade de dados gerados e com as regulações que ainda estão sendo implementadas, além da privacidade dos usuários.

Essas abordagens, embora complementares, ainda enfrentam limitações significativas quanto à aplicabilidade prática, fiscalização e compatibilidade com os princípios éticos e jurídicos que regem a atividade de inteligência e a segurança informacional.

Desafios éticos e legais

A crescente adoção da inteligência artificial em diferentes setores da sociedade evidência que as discussões sobre ética não podem se restringir ao desempenho técnico dos sistemas, devendo considerar também seus impactos sociais, jurídicos e políticos. Nesse contexto, Floridi e Cowls (2019) propõem um conjunto de princípios éticos fundamentais para orientar o desenvolvimento e a utilização da IA, baseado na beneficência, na não maleficência, na autonomia, na justiça e na explicabilidade. Esses princípios buscam assegurar que os sistemas inteligentes promovam benefícios à sociedade, minimizem danos, respeitem os direitos dos indivíduos e possibilitem que suas decisões sejam compreensíveis e justificáveis.

O trabalho de Watcher *et al*. (2016) segue na mesma linha, apresentando a existência de no mínimo sete tipos de problemas éticos que devem ser considerados fundamentais quando se analisa a criação, o funcionamento e o uso dos algoritmos, sendo a falibilidade, opacidade, viés, discriminação, autonomia, privacidade e responsabilidade.

A falibilidade contempla que todo sistema de AI e algoritmos são passíveis de erros, seja por falhas na coleta de dados, limitações nos modelos mate-

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

máticos, dificuldade em trabalhar com dados adicionados pelos usuários ou imprecisões nos processos de treinamento. Ela define que decisões automatizadas não podem ser consideradas infalíveis, e mecanismos de auditoria e revisão humana são essenciais para reduzir impactos negativos.

A opacidade aborda os algoritmos que são considerados 'caixas-pretas', ou seja, não permitem compreender como as decisões gerais foram geradas, é o caso das redes neurais e *deep learning* (aprendizado profundo). Assim, a opacidade compromete a transparência e dificulta a responsabilização, tornando necessário o desenvolvimento de ferramentas que permitam interpretar e explicar os processos internos de sistemas complexos, como é o caso da IA explicável. Complementando essa perspectiva, Dignum (2019) defende que a Inteligência Artificial Responsável (*Responsible Artificial Intelligence*) deve ser desenvolvida de forma legal, ética e tecnicamente robusta, considerando não apenas os aspectos tecnológicos, mas também os valores e objetivos da sociedade. Para a autora, a governança da IA envolve a definição de mecanismos de supervisão humana, prestação de contas (accountability), transparência e gestão contínua dos riscos, permitindo que organizações e desenvolvedores assumam responsabilidade pelos impactos decorrentes da utilização de sistemas automatizados.

Os vieses surgem quando os dados utilizados para treinamento dos algoritmos refletem desigualdades ou preconceitos existentes na sociedade e não necessariamente eles são englobados de forma proposital nos dados, mas de qualquer forma, tendem a levar a resultados enviesados que reproduzem ou amplificam injustiças sociais, econômicas ou culturais. De forma complementar, O'Neil (2016) evidência que esses modelos podem reproduzir e amplificar vieses presentes nos dados utilizados em seu treinamento, produzindo efeitos discriminatórios em áreas como concessão de crédito, educação, mercado de trabalho e justiça criminal. Esses estudos demonstram que a neutralidade algorítmica não pode ser presumida, reforçando a importância da adoção de princípios éticos e mecanismos de governança capazes de identificar, monitorar e mitigar vieses ao longo de todo o ciclo de vida dos sistemas de IA.

A discriminação ocorre quando decisões automatizadas prejudicam grupos específicos, seja por características étnicas, gênero ou idade. Sistemas mal projetados podem gerar recusa de crédito, erros em reconhecimento facial ou desigualdades em serviços públicos, evidenciando a necessidade de governança e a necessidade de revisão das decisões por seres humanos.

Quanto a autonomia, ela refere-se à capacidade de sistemas de IA realizarem

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

decisões sem intervenção humana direta, e embora aumente sua eficiência, também apresenta riscos, pois pode levar à tomada de decisão sem supervisão adequada, exigindo limites claros e políticas de responsabilidade.

A privacidade trata do uso excessivo de dados pessoais em algoritmos e IA, o que pode comprometer a privacidade dos indivíduos, e assim, garantir que informações sensíveis sejam protegidas e que haja consentimento informado é fundamental para manter a confiança dos usuários e atender à normas legais, como a LGPD (Lei Geral de Proteção de Dados) e a GDPR (Regulamento Geral sobre a Proteção de Dados).

Por fim, a responsabilidade envolve identificar e atribuir obrigações legais e éticas aos atores que projetam, implementam e utilizam sistemas de IA sendo essencial que existam mecanismos claros de prestação de contas, para que erros ou impactos negativos possam ser corrigidos e reparados de forma justa. Nesse contexto há uma grande discussão sobre quem deveria ser responsabilizado em caso de falhas, programadores, analistas, a própria empresa, entre outros, sendo uma discussão que ainda não teve desfecho e que tende a evoluir ao longo dos próximos anos.

Exemplos práticos de problemas enfrentados pelas empresas nesse cenário incluem a Amazon, a Apple e a Clearview AI. A Amazon, em 2014 lançou um algoritmo para agilizar a seleção de currículos, porém, currículos que continham a palavra "feminino" ou "mulher" eram penalizados em detrimento de currículos de homens, isso porque a área de tecnologia é majoritariamente masculina, fazendo com que o modelo não tivesse um comportamento neutro ao analisar e selecionar os currículos. A Apple, através do Apple Card estava concedendo um crédito menor para mulheres, mesmo quando elas tinham o mesmo perfil financeiro que o de homens.

Já a Clearview IA, uma empresa americana, foi multada em €30,5 Milhões pela Autoridade de Proteção de Dados Holandesa (DPA) por coletar de forma ilegal mais de 30 bilhões de imagens de redes sociais, inclusive de crianças, para criar uma base de dados global com o objetivo de permitir que autoridades fizessem busca por imagens.

Gillespie (2018) menciona que cada vez as plataformas estão sendo chamadas a responsabilidade de regular e bloquear determinados conteúdos, filtrando vídeos, mensagens e fotos que possam conter conteúdos considerados inaceitáveis, como desinformação, discurso de ódio ou conteúdo preconceituoso. Além disso, a pressão social também tem exigido soluções

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

técnicas para esses problemas, conforme Katzenbach e Ulbricht (2019). Porém conforme mencionado pelo autor, a linha entre a pesquisa e a otimização dos algoritmos de governança é muito tênue e implica em diversos fatores sociais, como a liberdade de expressão e neutralidade da rede, conforme o Marco Legal da Internet.

No contexto brasileiro, o trabalho de Pereira Filho e Lima (2025) apresenta um estudo da governança algorítmica voltada ao uso da IA nas políticas e administração pública, mencionando algumas ferramentas de IA que inclusive já estão sendo usadas, como a "ALICE", "Socrátes", "Bem-te-vi" e o "PIÁ". Os autores também ressaltam a importância na transparência e na ética voltada ao processo decisório e por fim, sugerem a criação de um framework jurídico específico para a administração pública, com foco na transparência algorítmica e no controle social.

Desafios quanto a eliminação dos vieses dos dados

Ao longo dos últimos anos foram criadas diversas iniciativas focadas no desenvolvimento de dados sem vieses, visando mitigar discriminações em sistemas de inteligência artificial, eles vão desde o desenvolvimento de ferramentas, abordagens acadêmicas, iniciativas governamentais e corporativas até mesmo as práticas de governança de dados.

Quanto as ferramentas, menciona-se a AI Fairness 360 (AIF360) criada pela IBM, a FairTest e a D-Bias. A AI Fairness 360 é um conjunto de ferramentas de código aberto que oferecem métricas de justiça e algoritmos para detectar e mitigar vieses em dados e modelos de IA (Bellamy 2018). Já a FairTest é uma metodologia que busca descobrir associações não justificadas em aplicações baseadas em dados, ajudando desenvolvedores a identificar e corrigir tratamentos discriminatórios, de acordo com Deutsch (2015). Por fim o sistema D-Bias é um sistema interativo que utiliza modelos causais para auditar e mitigar vieses sociais em conjuntos de dados tabulares, permitindo ajustes baseados em métricas de justiça e distorções de dados. O uso do modelo causal permite que os usuários identifiquem relações causais injustas, como viés contra grupos específicos (por exemplo, mulheres ou mulheres negras), e intervenham para mitigar os mesmos. Os testes realizados demonstraram que a ferramenta reduz os vieses em comparação com abordagens automatizadas, mantendo a integridade dos dados e melhorando a confiança, interpretabilidade e responsabilidade no processo decisório (Mishra 2022).

Com relação a conjuntos de dados criados de forma mais inclusiva, a empresa

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

Meta criou o conjunto de dados Casual Conversation v2, que contém mais de 25 mil vídeos de 5 mil pessoas em 7 países, incluindo informações sobre idade, gênero, raça e deficiências. Esses dados são úteis para treinamento de sistemas de reconhecimento de voz e facial. A empresa tem como expectativa, conforme Austin (2023), que outros programadores utilizem o conjunto para testar a precisão de seus sistemas em uma diversidade de pessoas, visando reduzir a discriminação automatizada.

Dentro das abordagens acadêmicas e técnicas tem-se a iniciativa FairVis e a Data Governance Models for Algorithms. A FairVis é um sistema de análise visual que auxilia na descoberta de vieses interseccionais (preconceitos ou discriminações que surgem da interseção de múltiplas identidades sociais de uma pessoa, como gênero, raça, idade, classe social, orientação sexual, deficiência, entre outros) em modelos de aprendizagem de máquina, o que permite que os usuários procurem deixar os modelos mais equalitários (Cabrera *et al*. 2019). Por fim, os Data Governance Models for Algorithms são modelos de governança de dados pessoais que abordam a discriminação algorítmica, propondo práticas para garantir a justiça nos sistemas automatizados (Iwan-Sojka 2025).

O Quadro 1 evidencia que as iniciativas analisadas apresentam avanços importantes na mitigação de vieses algorítmicos, porém concentram-se predominantemente em aspectos técnicos, como métricas de avaliação, tratamento dos dados e detecção de discriminações. Em contrapartida, observa-se que limitações relacionadas à necessidade de conhecimento especializado, à dependência de intervenção humana e à ausência de mecanismos automáticos de auditoria e responsabilização permanecem recorrentes entre as diferentes soluções. Sob a perspectiva da Inteligência Artificial Responsável, esses resultados indicam que a redução de vieses não depende exclusivamente de ferramentas computacionais, mas também da adoção de práticas de governança que incorporem transparência, supervisão humana e prestação de contas ao longo de todo o ciclo de vida dos sistemas (Dignum 2019).

Essa análise também corrobora a perspectiva de Yeung (2018), segundo a qual a governança algorítmica deve ultrapassar a dimensão técnica dos algoritmos e considerar mecanismos institucionais capazes de garantir auditabilidade e responsabilização. Embora iniciativas como AI Fairness 360, FairTest e D-BIAS contribuam para a identificação e mitigação de vieses, nenhuma delas oferece, isoladamente, uma solução abrangente para os desafios éticos associados à inteligência artificial. Esse resultado reforça a necessidade de combinar ferramentas técnicas com políticas organizacionais e estruturas regulatórias,

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

conforme defendido por Floridi e Cowls (2019), de modo a promover sistemas mais transparentes, justos e confiáveis.

Quadro 1 - Pontos positivos e negativos das iniciativas voltadas a eliminação de vieses nos dados

Iniciativa / Ferramenta	Pontos Positivos	Pontos Negativos
AI Fairness 360	Ampla variedade de métricas e algoritmos, documentação extensa, possui código aberto e facilita a transição para indústria.	Requer conhecimento técnico, implementação complexa em grandes bases de dados e não cobre todos tipos de vieses.
FairTest	Permite detectar discriminação contra subgrupos e ajuda a tratar vieses como "bugs" que podem ser corrigidos.	Depende de exploração manual.
D-BIAS	Permite intervenção humana, mantém integridade dos dados e reduz os vieses de forma controlada.	Complexidade de uso e depende da modelagem causal correta
Casual Conversations v2	Possui grande diversidade demográfica, o que ajuda a reduzir discriminação automatizada.	Serve apenas para treinamento/ teste, pois não corrige os algoritmos automaticamente, e ainda podem existir vieses residuais.
FairVis	Facilita identificação de vieses complexos através de uma abordagem visual intuitiva	Requer interpretação humana e não aplica correções automáticas.
Data Governance Models for Algorithms	Promove decisões inclusivas, abrangendo múltiplas dimensões sociais e culturais.	Eficácia depende de fiscalização e adesão e não atua diretamente em algoritmos existentes.

Fonte: elaboração própria

Por fim, observa-se que, apesar das diferentes abordagens adotadas, as limitações identificadas concentram-se em três aspectos recorrentes: elevada dependência de especialistas, baixa automação da mitigação de vieses e ausência de mecanismos abrangentes de governança e responsabilização, conforme Quadro 2.

Quadro 2 – Agrupamento das limitações das abordagens

Categoria	Ferramentas
Dependência de especialista	AIF360, D-BIAS, FairVis
Não corrige automaticamente	FairTest, Casual Conversations, FairVis
Não resolve governança	praticamente todas

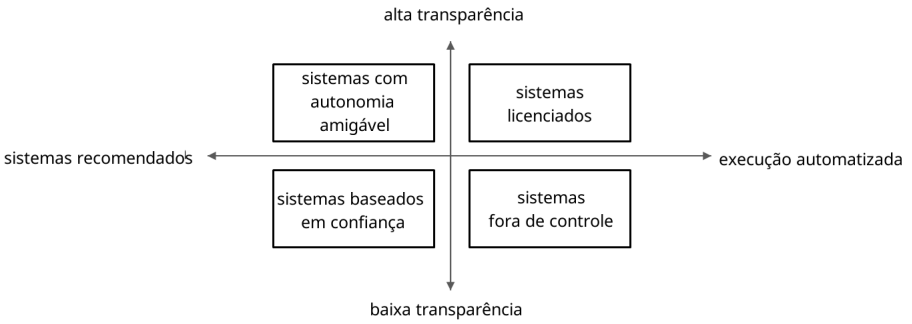
Fonte: Elaboração própria

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

Iniciativas em governança algorítmica

Tem-se discutido modelos de governança algorítmica, como o demonstrado na Figura 3, onde o grau de automação dos sistemas permite a tomada de decisão humana, confirmando ou não o resultado trazido pelo sistema, mas permitindo que a revisão humana altere esse resultado. Assim, conforme os autores Issar e Aneesh (2021), o ideal é uma combinação das duas dimensões e características, proporcionando sistemas com alta transparência e permitindo que humanos possam tomar decisões com base nos resultados trazidos pelos algoritmos.

Figura 3 - Tipos de sistemas de governança algorítmica



Fonte: Issar e Aneesh (2021)

Outra solução, apresentada por Lovatto (2024), é a explicabilidade em algoritmos de Inteligência Artificial (XAI), que se refere à capacidade de um sistema de IA de fornecer explicações claras e compreensíveis sobre como ele chegou às decisões apresentadas, permitindo que programadores, auditores e usuários em geral compreendam cada um dos fluxos e etapas por trás das decisões algorítmicas.

Esse tipo de solução deve estar alinhado com as políticas de governança da empresa, englobando o uso responsável de dados, a prevenção de discriminação e a proteção de direitos individuais e à privacidade, conforme Lovatto (2024). Além disso, deve ser uma entre outras práticas, como o monitoramento e a auditoria contínua.

Quanto as regulamentações sobre o tema, considera-se como a mais completa a AI Act da União Europeia, aprovada em 2023 e com entrada em vigor prevista para 2027, sendo que seu objetivo é não somente regular a IA, mas também promover seu uso responsável através de uma abordagem baseada em risco (Lovatto 2024). Para isso, os sistemas são classificados em 4 cate-

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

gorias: risco mínimo, risco limitado, alto risco e risco inaceitável (proibidas), com diferentes medidas a serem adotadas conforme o risco identificado.

Quanto ao Brasil, o trabalho de Sandrini e Pereira (2025) apresenta 11 projetos que já foram apresentados ao plenário desde 2019, e dessas, somente um prosperou, o PL nº 2.338/2023, as demais, em geral limitavam-se a versar sobre as diretrizes e princípios da IA basicamente reproduzindo as recomendações da OCDE.

Aprovou-se então, no Senado Federal a proposta de Lei 2.338/2023, conhecida como Marco Legal da Inteligência Artificial no Brasil (Brasil 2023), porém embora esteja em debate, ainda precisa ser aprovada pela Câmara dos Deputados e pelo Presidente para que entre em vigor. Ela é baseada na AI Act da União Europeia mas já sofreu 145 emendas, e tem como objetivo estabelecer princípios, direitos e deveres para o desenvolvimento e uso responsável da IA no país, inclusive com a proposta de criação de uma agência nacional de inteligência artificial, buscando criar um ambiente regulatório que promova a inovação tecnológica com segurança jurídica, assegurando a proteção de direitos fundamentais, a transparência, a responsabilização e a prevenção de riscos decorrentes de sistemas automatizados. Além disso, propõe diretrizes para a governança algorítmica, incentivo à pesquisa, supervisão ética e mecanismos de fiscalização, de modo a equilibrar o avanço tecnológico com o respeito à dignidade humana e à privacidade.

Outro projeto que tramita em paralelo com o anterior é o PL 210/2024 (Brasil 2024c), que dispõe sobre os princípios para o uso de tecnologia de inteligência artificial no Brasil, definindo que sistemas, programas ou processos que usam IA sejam regulados por normas que assegurem segurança, efetividade, proteção contra discriminação e garantia de privacidade e direitos dos indivíduos.

No entanto, as discussões nacionais não se restringem apenas a esfera federal, visto que diversos estados têm desenvolvido suas próprias iniciativas, refletindo o alcance do tema na administração pública (Sandrini e Pereira 2025). Em nível estadual, destacam-se seis estados com iniciativas de regulamentação ou Projetos de Lei sobre Governança de IA, conforme apresentado no Quadro 3, e observa-se, de forma geral, que a proteção dos direitos fundamentais, a privacidade e a não discriminação são temas recorrentes

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

entre as legislações e as iniciativas propostas.

Quadro 3 - Iniciativas de Governança de IA em Nível Estadual

Paraná	Lei nº 22.324/2025	Guiar a implementação ética e responsável da IA.	Implementação de um Plano Estratégico de Diretrizes.
Goiás	Lei Comp. nº 205/2025	Estimular o avanço tecnológico e a utilização segura da IA.	Estabelecimento de uma Política Estadual de Incentivo à Inovação.
Rio Grande do Sul	PL nº 179/2025	Estabelecer uma Política Estadual de Incentivo a respeito da inovação em IA.	Priorização do estímulo à inovação em IA.

Fonte: Sandrini e Pereira (2025)

A análise do Quadro 3 evidencia que as iniciativas estaduais apresentam uma convergência quanto à preocupação em promover o uso ético e responsável da inteligência artificial, especialmente por meio da proteção de direitos fundamentais, da supervisão humana e da definição de diretrizes para sua utilização no setor público. Entretanto, observa-se que a maior parte das normas possui caráter principiológico, estabelecendo objetivos gerais e orientações para o desenvolvimento da IA, mas apresentando poucos mecanismos concretos de fiscalização, auditoria ou responsabilização em casos de descumprimento.

Há ainda iniciativas em nível municipal, com 4 capitais brasileiras promovendo iniciativas voltadas ao tema através de leis (Curitiba), projetos de leis (Rio de Janeiro e Goiânia) e instrução normativa (Porto Alegre). Curitiba foi pioneira na criação de uma lei municipal específica através da Lei 16.321/2024 (Brasil 2024a), cujo objetivo é o de regulamentar o uso da IA na administração pública através da criação de uma estrutura de gestão.

O Rio de Janeiro, através da PL nº 2.970/2024 (Brasil 2024b), busca orientar o uso responsável de IA nos órgãos públicos e Goiânia, através da PL nº 240/2023 e 37/2025 procura criar diretrizes para uso, desenvolvimento e contratação de IA pela prefeitura, tendo como foco nas regras para contratação desses sistemas.

Por fim, Porto Alegre, através da Instrução Normativa PGM nº 003/2025 (Brasil 2025) busca regulamentar o uso setorial de IA no âmbito da Procuradoria-Geral do Município.

Assim, enquanto a regulamentação federal não é sancionada, observa-se uma postura proativa de diversos estados e municípios na criação de nor-

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

mas voltadas à utilização da inteligência artificial. Esse cenário demonstra o reconhecimento da importância da governança algorítmica para orientar o desenvolvimento e o uso responsável da IA, em consonância com os princípios de transparência, supervisão humana e responsabilização discutidos por Dignum (2019) e Floridi e Cowls (2019). Entretanto, a diversidade de iniciativas evidencia a ausência de um modelo regulatório uniforme, podendo demandar futura revisão com a legislação federal para assegurar maior segurança jurídica e consistência na proteção dos direitos fundamentais.

Outro aspecto observado é a diversidade de estratégias adotadas pelos estados. Enquanto Ceará, Alagoas e Paraná priorizam princípios relacionados à ética, transparência e supervisão humana, Amazonas, Goiás e Rio Grande do Sul concentram seus esforços no incentivo à inovação e na estruturação de políticas públicas voltadas ao desenvolvimento da inteligência artificial. Essa diversidade demonstra que, embora exista consenso quanto à importância da regulamentação, ainda não há uniformidade na definição dos instrumentos de governança algorítmica.

Quanto aos discursos de ódio e *fake news*, diversos países têm proposto formas de controlar o conteúdo divulgado em redes sociais, como no Brasil através da PL 2.630/2020 (Brasil 2020), conhecido como Lei das *Fake News*, cujo objetivo é o de combater a disseminação de informações falsas e enganosas na internet, estabelecendo responsabilidades para plataformas digitais e usuários. Entre as medidas propostas estão a obrigatoriedade de identificação de usuários, a criação de mecanismos de verificação de conteúdo e a responsabilização das plataformas por conteúdos prejudiciais. O projeto foi aprovado pelo Senado em 2020 e encaminhado à Câmara dos Deputados, onde segue em análise e com grande polarização do público devido a seu pequeno limiar entre censura e controle de discurso de ódio.

Com relação à governança algorítmica de fato, boa parte das plataformas digitais tem atualizado seus termos de uso e privacidade coibindo a divulgação de notícias falsas e discursos de ódio. Além dessas ações corporativas, surgiram diversas iniciativas globais que buscam promover o uso ético e responsável da tecnologia, como a Liga Anti-Difamação (ADL) e a Algorithmic Justice League. A Liga Anti-Difamação (ADL) atua no combate ao antissemitismo e na gestão de informações que auxiliam autoridades policiais e comunidades na prevenção de ameaças extremistas de diferentes naturezas. Já a Algorithmic Justice League tem se destacado como um movimento voltado à promoção de uma inteligência artificial equitativa e transparente, buscando reduzir vieses algorítmicos e incentivar práticas de desenvolvimento

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

mais justas e responsáveis.

Além dessas, outras iniciativas internacionais têm buscado estabelecer diretrizes, regulamentações e princípios éticos para o desenvolvimento e uso responsável da inteligência artificial e da governança algorítmica, como a Unesco e a OCDE. A UNESCO lançou diretrizes globais em 2021 que priorizam direitos humanos, equidade, transparência e responsabilidade na IA, enquanto a OCDE publicou princípios voltados à inovação responsável, robustez, supervisão humana e minimização de vieses.

Países como Canadá, Estados Unidos, Japão e Coreia do Sul também têm criado códigos de conduta, estratégias nacionais de IA e órgãos de supervisão para orientar empresas e órgãos públicos no uso seguro e ético da tecnologia. Essas iniciativas refletem uma preocupação global em equilibrar inovação tecnológica, segurança, privacidade e proteção de direitos humanos.

O Quadro 4 apresenta as iniciativas discutidas, bem como alguns pontos positivos e negativos em cada uma delas. Já o Quadro 5 consolida os modelos de governança, iniciativas práticas e instrumentos de regulação, complementando a análise através da perspectiva por abordagem, objetivos e limitações.

Quadro 4 - Soluções encontradas que buscam mitigar os problemas apresentados no trabalho.

Iniciativas	Pontos Fortes	Pontos Fracos
AI Act da União Européia	Regulação robusta não somente da IA mas do seu uso responsável, envolvendo a classificação de seu uso conforme risco.	Pode impor altos custos de conformidade, dificultando a inovação e a competitividade de pequenas empresas e startups no setor de inteligência artificial.
Marco Legal da Inteligência Artificial no Brasil, Lei 2.338/2023	Busca promover a segurança jurídica, incentiva a inovação e estabelece princípios éticos e de transparência para o uso responsável da inteligência artificial.	Texto genérico, com lacunas sobre fiscalização, aplicação prática e sanções, podendo gerar insegurança regulatória e dificuldades de implementação. Sem previsão de entrada em vigor.
PL 210/2024	Estabelece princípios éticos e jurídicos para o uso responsável da inteligência artificial no Brasil.	Também genérico e pode conflitar com a Lei 2.338/2023 pois retoma e simplifica princípios já abordados, o que pode gerar sobreposição normativa e insegurança jurídica, além de atrapalhar o andamento do debate que já está mais consolidado na outra regulamentação.
PL 2630/2020	Visa combater a desinformação e responsabiliza plataformas digitais, promovendo transparência e proteção a usuários vulneráveis.	A discussão a respeito desse controle se relaciona a censura e ao risco de privacidade, pois não existe uma unanimidade quanto a quem ou como iria fiscalizar essas mensagens e se esse órgão não agiria com viés político.

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

Plataformas digitais	Buscam reduzir a propagação de notícias falsas e discursos de ódio, promovendo maior segurança e confiabilidade na internet.	Podem gerar censura e decisões arbitrárias e alinhados a gestão da empresa, além de dependerem da fiscalização interna das próprias empresas.
Liga Anti Difamação	Auxilia na proteção de comunidades e autoridades contra ameaças extremistas, promovendo educação e conscientização sobre ódio e discriminação.	Sua atuação depende de financiamento e parcerias externas, podendo ter alcance limitado e foco restrito a determinadas comunidades.
Algorithmic Justice League	Busca promover o desenvolvimento de uma IA ética, equitativa e responsável, reduzindo vieses e aumentando a transparência nos algoritmos.	Sua influência prática sobre empresas e governos é limitada, e muitas recomendações podem ser de difícil implementação em larga escala.
Unesco	Procura promover princípios universais de direitos humanos e equidade na IA.	Sua implementação depende da adesão voluntária dos países, sem mecanismos de fiscalização obrigatórios.
OCDE	Tem como objetivo incentivar a inovação responsável e mitigação de vieses algorítmicos.	São apenas recomendações, não possuindo força legal vinculante.

Fonte: Elaboração própria

Quadro 5 - Perspectiva dos modelos de governança algorítmica por abordagem, objetivos e limitações

Abordagem / Iniciativa	Descrição	Objetivo principal	Limitações / Desafios
Autorregulação das plataformas digitais	Políticas internas, termos de uso e moderação de conteúdos	Reduzir fake news, discursos de ódio e promover segurança online	Dependente da vontade da empresa, risco de censura e decisões arbitrárias e nem sempre isentas de vieses
Regulação estatal (PL 2630/2020, PL 2338/2023)	Legislação específica para IA e desinformação	Estabelecer regras obrigatórias, responsabilização e transparência	Aplicação complexa, custos de conformidade, risco de fragmentação normativa
Normas internacionais (UNESCO, OCDE, AI Act)	Diretrizes globais ou regionais para IA ética e segura	Garantir uso responsável, equidade e proteção de direitos humanos	Não são vinculantes globalmente (UNESCO/ OCDE), custos de conformidade (UE), grande dificuldade de harmonização internacional
Movimentos civis e organizações (AJL, ADL)	Iniciativas de sociedade civil e ONGs	Reduzir vieses, promover justiça algorítmica e proteção de comunidades	Influência prática limitada, recomendações de difícil implementação em larga escala
Auditorias e certificações de IA	Avaliação independente de sistemas de IA	Validar transparência, rastreabilidade e ética algorítmica	Custos altos e padronização ainda em desenvolvimento e sem framework definido
Modelos de governança algorítmica corporativa	Comitês internos, políticas de compliance e ética	Garantir decisões automatizadas mais responsáveis	Muitas vezes superficiais, sem fiscalização externa

Fonte: Elaboração própria

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

A análise dos Quadros 4 e 5 evidencia que a governança algorítmica ainda se encontra em diferentes estágios de maturidade entre países e organizações. Enquanto algumas iniciativas estabelecem mecanismos regulatórios mais estruturados, outras permanecem restritas à formulação de princípios ou recomendações de caráter voluntário, refletindo diferentes estratégias para equilibrar inovação tecnológica, proteção de direitos fundamentais e desenvolvimento econômico.

Esse cenário reforça o desafio da governança algorítmica em equilibrar interesses relacionados à segurança, vigilância, privacidade e liberdades civis. Conforme discutido por Floridi e Cowls (2019), o desenvolvimento ético da inteligência artificial depende da conciliação entre inovação tecnológica e proteção dos direitos fundamentais, especialmente em aplicações relacionadas ao combate à desinformação e às ameaças à segurança nacional.

Observa-se uma convergência entre as diferentes iniciativas quanto à incorporação de princípios relacionados à transparência, responsabilidade, proteção de dados e respeito aos direitos humanos. Entretanto, a análise também evidencia uma lacuna recorrente: a maioria das iniciativas concentra-se na definição de princípios gerais, apresentando menor detalhamento sobre mecanismos de auditoria independente, monitoramento contínuo e responsabilização dos agentes envolvidos. Esse resultado corrobora a perspectiva de Yeung (2018), segundo a qual a governança algorítmica depende não apenas da existência de normas, mas da capacidade institucional de fiscalizar sua efetiva implementação.

Os resultados obtidos indicam que uma governança algorítmica efetiva exige a integração entre instrumentos regulatórios, mecanismos técnicos de mitigação de vieses e estruturas organizacionais de supervisão. Nesse contexto, tecnologias de XAI, auditorias independentes e modelos corporativos de governança representam instrumentos complementares para aumentar a transparência e a confiabilidade dos sistemas inteligentes. Além disso, considerando o caráter transnacional da desinformação, das ameaças cibernéticas e do fluxo internacional de dados, observa-se que a cooperação entre países tende a desempenhar papel estratégico para a harmonização/ padronização de normas e o fortalecimento da governança da inteligência artificial.

Por fim, a classificação apresentada no Quadro 5 permite observar que a governança algorítmica não depende de um único instrumento regulatório, mas da atuação complementar de diferentes mecanismos institucionais. A autorregulação das plataformas, embora contribua para respostas rápi-

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

das, apresenta limitações relacionadas à transparência e aos conflitos de interesse. As regulamentações estatais conferem maior segurança jurídica, mas enfrentam desafios de implementação e atualização diante da rápida evolução tecnológica. Por sua vez, normas internacionais e organizações da sociedade civil exercem papel relevante na consolidação de princípios éticos, ainda que sua efetividade dependa da adesão voluntária e da capacidade de influência sobre governos e empresas. Finalmente, auditorias independentes e modelos corporativos de governança representam mecanismos promissores para operacionalizar princípios éticos na prática, embora ainda careçam de padronização e mecanismos consolidados de fiscalização.

Conclusão

O papel dos sistemas de informação na sociedade contemporânea é crescente e estrutural, especialmente diante da expansão da coleta e do processamento de dados para finalidades como recomendação de conteúdo, previsão de comportamentos, prevenção de crimes e análise de mercado. Nesse contexto, consolida-se um cenário de crescente dependência de sistemas algorítmicos na tomada de decisão, o que reforça a centralidade da governança algorítmica.

Entretanto, observa-se que o uso intensivo desses sistemas frequentemente incorpora vieses presentes nos dados e nos modelos, podendo gerar impactos negativos relevantes, como discriminação em concessão de crédito, falhas em sistemas de reconhecimento facial e decisões automatizadas opacas. Embora legislações como a LGPD e o GDPR estabeleçam avanços importantes na proteção de dados e na limitação da coleta e tratamento de informações pessoais, ainda persistem desafios relacionados à fiscalização, auditoria e efetiva responsabilização dos agentes envolvidos.

Do ponto de vista técnico, identificam-se esforços contínuos voltados à mitigação de vieses, melhoria da qualidade dos dados e desenvolvimento de modelos mais robustos e diversificados. Contudo, conforme discutido na literatura de governança algorítmica, tais soluções não são suficientes quando analisadas isoladamente, uma vez que a mitigação de riscos exige também mecanismos institucionais de supervisão e responsabilização ao longo de todo o ciclo de vida dos sistemas (Yeung 2018; Dignum 2019).

Nesse sentido, a análise realizada evidencia que a governança de algoritmos no contexto da inteligência deve operar de forma multidimensional, integrando aspectos técnicos, legais e organizacionais, com ênfase na transparência, auditabilidade, mitigação de vieses e responsabilização humana pelas deci-

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

sões automatizadas. Assim, mais do que a criação de normas isoladas ou ferramentas técnicas, faz-se necessária a articulação entre diferentes níveis de governança para lidar com os impactos crescentes da automação decisória.

Por fim, este estudo contribui para a literatura ao sistematizar e analisar iniciativas de governança algorítmica, identificando padrões, limitações e lacunas comuns entre abordagens técnicas, regulatórias e institucionais. Os resultados indicam predominância de soluções principiológicas e desafios persistentes relacionados à auditoria, transparência e responsabilização, reforçando a necessidade de uma governança multidimensional que integre aspectos éticos, legais e técnicos.

Como trabalhos futuros tem-se a necessidade de ampliação da análise para outros fenômenos associados à governança algorítmica, como moderação de conteúdo, desinformação e sistemas de vigilância, de modo a aprofundar a compreensão dos mecanismos de controle, mitigação de riscos e proteção de direitos fundamentais no ambiente digital.

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

Referências

Aneesh, Anup. 2006. Virtual Migration: The Programming of Globalization. Durham: Duke University Press.

Austin Jr., Roy; Fried, Ina. 2023. "Meta's new yardstick for testing algorithmic bias". Axios, 9 mar. 2023. https://www.axios.com/2023/03/09/algorithmic-bias-meta-dataset-face-recognition

Bellamy, Rachel K. E., et al. 2018. "AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias". https://arxiv.org/abs/1810.01943

Brasil. 2020. Projeto de Lei nº 2.630/2020. https://www25.senado.leg.br/web/atividade/materias/-/materia/141944.

Brasil. 2023. Projeto de Lei n° 2.338/2023. https:// www25.senado.leg.br/web/atividade/materias/-/materia/157233

Brasil. 2024a. Lei nº 16.321/2024. https://leismunicipais.com.br/a/pr/c/curitiba/lei-ordinaria/2024/1633/16321

Brasil. 2024b. PL nº 2.970/2024. https://mail.camara.rj.gov.br/APL/Legislativos/scpro.nsf/ab87ae0e15e7dddd0325863200569395/03258c16006f052703258ae6006571b4?OpenDocument

Brasil. 2024c. Projeto de Lei nº 210. https://legis.senado.leg.br/sdleg-getter/documento?disposition=inline&dm=9543976&ts=1738768205705

Brasil. 2025. Instrução Normativa PGM nº 003/2025. https://dopaonlineupload.procempa.com.br/dopaonlineupload/5567_ce_20250228_executivo.pdf

Brisola, A.; Bezerra, A. C. 2018. "Desinformação e Circulação de Fake News: distinções, diagnóstico e reação". Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB – XIX. Ed. ANCIB, pp. 3316–3330. https://hdl.handle.net/20.500.11959/brapci/102819

Cabrera, Ángel Alexander, et al. 2019. "FairVis: Visual Analytics for Discovering Intersectional Bias in Machine Learning". https://arxiv.org/abs/1904.05419

Cooper, F. R. 2014. "Always Already Suspect: Revising Vulnerability Theory". North Carolina Law Review 93.

Doneda, Danilo, e Almeida, Virgílio A. F. (2016). "What Is Algorithm Governance?". *Internet Governance*, IEEE Computer Society, 1089-7801.

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

De Mello, F. C. O., e Scheinder, M. (2021). ‘‘Desinformação Digital em Rede e Competência Crítica em Informação’’. *International Review of Information Ethics*, Vol. 30.

Deutsch, William et al. (2015). ‘‘FairTest: Discovering Unwarranted Associations in Data-Driven Applications’’. Disponível em: https://arxiv.org/abs/1510.02377. Acesso em: 02 out. 2025.

Dignum, Virginia. 2019. Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Cham: Springer.

Errey-Hammond, M. 2024. Big Data, Emerging Technologies and Intelligence. Abingdon: Routledge.

Floridi, Luciano; Cowls, Josh. 2019. ‘‘A unified framework of five principles for AI in society’’. Harvard Data Science Review, Cambridge, v. 1, n. 1. https://doi.org/10.1162/99608f92.8cd550d1

Ford, C.; Vogel, K. 2025. Will AI Save Us?: The Role of Artificial Intelligence in Intelligence Analysis. In: US Intelligence Failure and Knowledge Creation: Improving Intelligence Analysis. Abingdon: Routledge.

Gillespie, Tarleton. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven: Yale University Press.

Issar, S., e Aneesh, A. (2021). ‘‘What is Algorithmic Governance?’’. Wiley.

Iwan-Sojka, D. (2025). ‘‘The inclusive data governance models for algorithms’’. *Information, Communication & Society*, v. 29, n. 1, p. 1–20. Disponível em: https://www.tandfonline.com/doi/full/10.1080/13600834.2024.2406668. Acesso em: 23 ago. 2025.

Junqueira, A. H., e Botelho-Francisco, R. (2021). ‘‘Raça: Dimensão Interseccional das Vulnerabilidades Digitais’’. *Contemporanea – Revista de Comunicação e Cultura*, v. 19, n. 3.

Katzenbach, C., e Ulbricht, L. (2019). ‘‘Algorithmic Governance’’. *Internet Policy Review – Journal on Internet Regulation*, v. 8, n. 4.

Lovatto, M. Betiele Aude. (2024). ‘‘Inteligência Artificial: Governança e Transparência?’’. *Revista Ibmec de Direito*, 1(1). Disponível em: https://ibmec.periodicoscientificos.com.br/index.php/cienciajuridica/article/view/245. Acesso em: 05 out. 2025.

Luna, F. (2009). ‘‘Elucidating the Concept of Vulnerability: Layers Not Labels’’. *The International Journal of Feminist Approaches to Bioethics (Spring)*, Vol. 2, No. 1.

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

Lyon, D. (2014). ‘‘Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique’’. *Big Data & Society*, 1(2). Disponível em: https://journals.sagepub.com/doi/10.1177/2053951714541861. Acesso em: 23 ago. 2025.

Malgieri, G., e Niklas, J. (2020). ‘‘Vulnerable Data Subjects’’. *Computer Law & Security Review*, 37. Elsevier.

Mishra, Neha et al. (2022). ‘‘D-BIAS: Human-in-the-Loop Debiasing for Tabular Data Using Causal Models’’. Disponível em: https://arxiv.org/abs/2208.05126. Acesso em: 05 out. 2025.

Müller-Birn, C., Herbsleb, J., e Dobusch, L. (2013). ‘‘Work-to-Rule: The Emergence of Algorithmic Governance in Wikipedia’’. *Proceedings of the 6th International Conference on Communities and Technologies*, 80–89. Disponível em: https://dl.acm.org/doi/10.1145/2482991.2482999. Acesso em: 23 ago. 2025.

O’Neil, Cathy. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown.

O’Reilly, T. (2013). ‘‘Open Data and Algorithmic Regulation’’. Em B. Goldstein & L. Dyson (eds.), *Beyond Transparency: Open Data and the Future of Civic Innovation* (pp. 289–300). San Francisco: Code for America Press.

Pereira Filho, Nivanildo, e Lima, Rogerio de Araujo. (2024). ‘‘Governança Algorítmica e Políticas Públicas: Desafios Éticos e Impactos da Inteligência Artificial na Tomada de Decisão Governamental’’. *RECIMA21 - Revista Científica Multidisciplinar*, v. 6, n. 1, p. e616051. DOI: 10.47820/recima21.v6i1.6051. Disponível em: https://recima21.com.br/recima21/article/view/6051. Acesso em: 05 out. 2025.

Posetti, J., e Bontcheva, K. (2020). ‘‘Desinfodemia: descifrando la desinformación sobre el COVID-19’’. Paris: UNESCO. Disponível em: https://en.unesco.org/sites/default/files/disinfodemic_deciphering_covid19_disinformation_es.pdf. Acesso em: 12 set. 2025.

Rouvroy, A., e Berns, T. (2013). ‘‘Gouvernementalité algorithmique et perspectives d’émancipation. Le disparate comme condition d’individuation par la relation?’’. *Réseaux*, n. 177, p. 163–196. Disponível em: https://www.cairn.info/revue-reseaux-2013-1-page-163.htm. Acesso em: 24 ago. 2025.

Sandrini, Camila Taciana, e Pereira, Camila Gourgues. (2025). ‘‘Governança de Inteligência Artificial no Brasil: Panorama Regulatório Atual’’. *Ciências Sociais Aplicadas*, v. 29, n. 149. Disponível em: https://doi.org/10.69849/revistaft/ni10202508072123. Acesso em: 05 out. 2025.

Revista Brasileira de Inteligência
2026 • nº 21 • e2026.21.291
ISSN 2595-4717

Stenslie, S. (ed.). 2022. Intelligence Analysis in the Digital Age. Abingdon: Routledge.

Volkoff, Vladimir. (2004). *Pequena História da Desinformação: Do Cavalo de Tróia à Internet*. Curitiba: Ed. Vila do Príncipe.

Yeung, Karen. 2018. "Algorithmic regulation: A critical interrogation". Regulation & Governance, Hoboken, v. 12, n. 4, p. 505–523. DOI: https://doi.org/10.1111/rego.12158

Wachter, S., e Mittelstadt, B. (2019). "A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI". *Columbia Business Law Review*.