Governança Algorítmica e Accountability Invisível: O Paradoxo do Controle na era da IA no Setor Público

Daniel Marques Trindade
Secretaria de Estado da Educação do Espírito Santo, Brasil
E-mail: dmtrindade@sedu.es.gov.br

Resumo

A incorporação crescente de sistemas de inteligência artificial na administração pública tem redefinido as fronteiras da tomada de decisão estatal. Ao mesmo tempo em que prometem eficiência, precisão e celeridade, os algoritmos introduzem uma lógica decisória que escapa dos mecanismos tradicionais de controle e transparência. Este artigo teórico investiga o surgimento do que se pode denominar "governança algorítmica invisível", marcada pela opacidade técnica, dificuldades de auditoria e despersonalização da responsabilidade institucional. Com base em revisão crítica da literatura e análise de experiências internacionais — como Reino Unido, Estônia e diretrizes da OCDE — o estudo identifica os riscos inerentes à adoção de IA sem salvaguardas adequadas: decisões automatizadas não rastreáveis, vieses sistêmicos e erosão da accountability democrática. Propõe-se, ao final, um conjunto de diretrizes públicas para garantir que a incorporação de IA na gestão pública seja acompanhada de mecanismos de explicabilidade algorítmica, supervisão humana qualificada e responsabilização institucional. A contribuição reside em oferecer um referencial crítico e propositivo para a construção de modelos de governança algorítmica compatíveis com os princípios do Estado Democrático de Direito.

Palavras-chave: Inteligência artificial; Governança algorítmica; Accountability pública; Decisão automatizada; Transparência no setor público.

1. Introdução

Nos últimos dez anos, órgãos governamentais passaram a adotar nas fases decisórias que antes dependiam unicamente de servidores públicos, sistemas que utilizam inteligência artificial (IA). Almejando ganhos em eficiência, redução de despesas e uniformização de procedimentos, esta ação animou os gestores; contudo, logo vieram à tona impactos colaterais que colocam em xeque princípios democráticos fundamentais; transparência, equidade e dever de prestar contas (Busuioc, 2020; BARREDO e ARRIETA et al., 2020). No contexto brasileiro, o Sistema de Seleção Aduaneira por Aprendizado de Máquina (SISAM) e o Analisador de Licitações e Editais (ALICE) ilustram, simultaneamente, progressos operacionais e conflitos constitucionais relacionados ao artigo 37 da CF/1988 e à Lei 9.784/1999 (Brasil, 1988; Brasil, 1999).

Nesse cenário, o presente trabalho examina os paradoxos que acompanham a inserção da IA nos processos de tomada de decisão na administração pública brasileira, focalizando três grupos de risco: (1) opacidade algorítmica, (2) vieses estruturais e discriminações automáticas e (3) diluição da responsabilidade institucional. O objetivo geral consiste em verificar de que forma esses riscos fragilizam práticas democráticas e, a partir dessa constatação, sugerir orientações de governança fundamentadas em explicabilidade, supervisão humana capacitada e

ambiente favorável a responsabilização. Os objetivos específicos deste estudo são:

- Levantar, no acervo acadêmico e em literatura, indícios de opacidade, enviesamentos e lacunas de accountability em órgãos da esfera federal.
- Examinar como os dispositivos normativos internos (CF/1988, Lei 9.784/1999, LGPD) respondem aos desafios introduzidos pela IA.
- Apontar experiências exitosas no exterior (Reino Unido, Canadá, Estônia, UE) e averiguar sua viabilidade de adaptação ao cenário brasileiro.
- Construir um modelo de governança algorítmica responsável voltado ao Poder Executivo da União.

Estruturalmente, o artigo abre com uma revisão teórica acerca dos três grupos de risco, avança para a seção de Resultados e Discussão, na qual se investigam exemplos no Brasil e fora dele, e conclui apresentando recomendações práticas embasadas tanto em experiências externas quanto nas exigências constitucionais brasileiras.

2. Metodologia

Este artigo foi concebido como um estudo qualitativo de tipo exploratório-descritivo, sustentado por procedimentos de pesquisa documental aliados à análise comparativa. Tal configuração metodológica é apontada pela literatura clássica como adequada quando a intenção é compreender fenômenos institucionais ainda em fase de consolidação (YIN, 2014; FLICK, 2023).

2.1 Estrutura epistemológica

- Pesquisa aplicada, considerando que se direciona à formulação norteadores de governança para entes públicos (GIL, 2019).
- Abordagem indutivo-analítica: parte-se da averiguação de casos empíricos (SISAM, ALICE, COMPAS, Reino Unido, Canadá, Estônia) para erigir proposições teóricas; essa lógica de “grounded propositions” é defendida por STRAUSS e CORBIN (2008).
- Perspectiva sócio-jurídica, combinando aportes da Administração Pública, Ciência de Dados e Direito, consonante ao modelo transdisciplinar proposto por NICHOLAOS (2017) em estudos sobre tecnologia e sociedade.

2.2 Critérios de validade e confiabilidade

Rastreabilidade documental: todas as referências contam com DOI ou URL permanente; nos casos em que isso não ocorre, podem ser localizadas com facilidade na web, assegurando a verificabilidade do material utilizado (MILES, HUBERMAN e SALDAÑA, 2018).

2.3 Limitações

A investigação apoia-se predominantemente em fontes secundárias e não contempla entrevistas com desenvolvedores privados de IA; tal recorte decorreu de restrições de tempo e de acesso, limitação reconhecida por CRESWELL (2013) em estudos de natureza exploratória. Pesquisas futuras poderão adotar métodos mistos para capturar as percepções dos atores diretamente impactados.

3. Resultados e Discussão

A implantação de ferramentas pautadas em IA na administração pública vem expondo contradições que tensionam alicerces democráticos tradicionais. Apesar de prometer ganhos de eficiência e maior objetividade, a automatização das deliberações administrativas pode originar perigos capazes de comprometer transparência, equidade e responsabilização institucional.

Esta seção explora os obstáculos mais recorrentes apontados pela literatura e por casos empíricos, agrupando-os em três grandes categorias de risco, com atenção especial ao panorama brasileiro e às práticas do Poder Executivo federal.

3.1 Opacidade algorítmica

O primeiro risco aponta à opacidade algorítmica, situação que se dá pela dificuldade de compreender e justificar escolhas automatizadas, um aspecto particularmente sensível no setor público, onde a transparência é princípio constitucional (BURRELL, 2016). O próprio BURRELL (2016) apresenta a noção de inteligibilidade fraca (weak intelligibility) e descreve três fontes distintas de opacidade: 1) opacidade intencional (corporate or state secrecy); 2) opacidade derivada do analfabetismo técnico (technical illiteracy); e 3) opacidade intrínseca aos algoritmos dotados de aprendizado de máquina (AM). Esta última se revela a mais desafiadora por se relacionar às próprias características técnicas dos sistemas contemporâneos de IA (BRASIL, Ministério da Saúde, 2023; MONTEIRO, 2021).

A dificuldade ganha contornos adicionais quando se percebe que muitos modelos de AM funcionam como “caixas-pretas”, impossibilitando até mesmo seus criadores de explicar plenamente a cadeia causal de decisão. Não é apenas um problema técnico, mas antes de uma questão epistemológica: a complexidade crescente desses sistemas enfraquece o modelo clássico de accountability baseado em explicações causais (BURRELL, 2016; BRASIL, Ministério da Saúde, 2023; MONTEIRO, 2021).

Um caso britânico ilustra bem as implicações práticas dessa opacidade. Em 2018, um sistema governamental automatizado foi descontinuado depois de inúmeras denúncias relativas a decisões injustas e obscuras. O episódio evidenciou a insuficiência dos instrumentos jurídicos vigentes para assegurar transparência em ambientes algorítmicos sofisticados. Embora a LGPD Britânica preveja o direito à explicabilidade, tal garantia mostrou-se incapaz de traduzir o raciocínio algorítmico de forma efetiva. Conforme WACHTER, MITTELSTADT e FLORIDI (2017), há um descompasso entre o direito à explicabilidade frequentemente invocado e aquilo que

o GDPR realmente assegura: apenas “informações significativas, porém adequadamente limitadas” sobre a lógica envolvida e sobre o alcance e as os desdobramentos de decisões realizadas por máquinas, de forma automatizada, o chamado right to be informed (BRASIL, Ministério da Saúde, 2023; MONTEIRO, 2021).

No Brasil, o SISAM (Sistema de Seleção Aduaneira por Aprendizado de Máquina), em operação na Receita Federal desde 2014, ilustra essa questão no âmbito do Executivo federal. Ainda que JAMBEIRO FILHO, apud KÖCHE, 2021, apud CUNHA, STAMILE e GRUPENMACHER (2024) ressaltem a preocupação do projeto com auditabilidade, incluindo logs detalhados do raciocínio, subsiste o fato de que a maioria dos usuários desconhece os parâmetros utilizados para a escolha de fiscalização. Nessa linha, defende-se uma transparência restrita quanto aos aspectos técnicos da modelagem, de modo a mater a eficácia fiscalizatória; de outra maneira, corre-se o risco de esvaziar a finalidade do processo, justificando uma exceção implícita à regra geral de publicidade prevista no § 1º do artigo 20 da LGPD (CUNHA, STAMILE e GRUPENMACHER, 2024; BRASIL, Ministério da Saúde, 2023; MONTEIRO, 2021).

Há, portanto, impactos constitucionais relevantes. O art. 37 da CF/88 impõe legalidade, impessoalidade, moralidade, publicidade e eficiência à administração pública. Se as decisões produzidas pelo SISAM não puderem ser devidamente motivadas, corre-se o risco de violar o princípio da motivação previsto na Lei 9.784/99, em que todo ato administrativo deve indicar os pressupostos de fato e de direito que o embasam (BRASIL, 1999).

> A incerteza acerca do funcionamento interno da IA e a possibilidade de erros ou vieses (Faúndez-Ugalde; Mellado-Silva; Aldunate-Lizana, 2020, p. 6) parecem colidir, à primeira vista, com as garantias constitucionais do contraditório, da ampla defesa e do devido processo legal. Decisões administrativas baseadas em caixas-pretas algorítmicas também desafiam a exigência de motivação dos atos (CUNHA; STAMILE; GRUPENMACHER, 2024, p. 6).

O dilema, no setor público nacional, emerge como tensão entre demandas operacionais e suas exigências legais no que concerne a transparência. CUNHA, STAMILE e GRUPENMACHER (2024) observam que, iniciativas como o SISAM justificam certa opacidade para preservar a efetividade fiscal. Todavia, o Supremo Tribunal Federal, no julgamento da ADI 6.649, advertiu que, diante das novas tecnologias da informação, a administração jamais pode comprometer princípios constitucionais fundamentais (STF, Rel. Min. Gilmar Mendes, 2022)[1].

3.2 Vieses sistêmicos e discriminação automatizada

O segundo grupo de riscos está na perpetuação e, muitas vezes, ao agravamento de preconceitos discriminatórios por meio de ferramentas automatizadas. Em vez de assegurar maior neutralidade, a IA generativa tem

[1] https://portal.stf.jus.br/processos/downloadPeca.asp?id=15358978491eext=.pdf

reproduzido, e até intensificado, desigualdades históricas associadas a atributos protegidos, como raça, gênero e condição socioeconômica.

Modelos generativos, não se baseiam em regras explícitas; eles produzem respostas mediante a modelagem estatística de vastos corpora. Cada saída deriva de cálculos de probabilidade condicionada: por meio de um prompt, o modelo estima, com base em distribuições internalizadas, qual palavra, sentença ou pixel provavelmente virá em seguida (RAN HE, JIE CAO, TIENIU TAN, 2025).

> Esses métodos capturam propriedades das distribuições de dados e erigem modelos estatísticos de forma explícita ou implícita. A estratégia explícita mais comum envolve modelos gráficos probabilísticos [(javascript:;)]; nesses gráficos, nós representam variáveis aleatórias que descrevem os dados e arestas refletem dependências probabilísticas. A geração de dados, então, pode ser vista como a inferência das partes desconhecidas do grafo, processo normalmente otimizado por algoritmos de maximização de verossimilhança, notadamente o Expectation–Maximization [(javascript:;)]. Além disso, tais modelos costumam pressupor a propriedade de Markov [(javascript:;)], segundo a qual o estado futuro depende exclusivamente do presente (RAN HE, JIE CAO, TIENIU TAN, 2025, p. 4).

Nos Estados Unidos, o software COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) evidencia esse problema. Empregado no sistema criminal visando averiguar o risco de reincidência, o COMPAS foi analisado pela ProPublica em 2016, que identificou vieses raciais substanciais. Ao confrontar as pontuações do programa com o comportamento real dos réus dois anos após a liberdade provisória, verificou-se que o algoritmo classificava falsamente réus negros como futuros infratores quase o dobro do que acontecia com brancos (LARSON et al., 2016): 45% dos réus negros que não reincidiram foram apontados como alto risco, contra 23% dos brancos em situação idêntica. Embora a taxa global de erro se mantivesse similar, a natureza dos equívocos diferia: réus negros eram superavaliados, acarretando fianças mais altas, detenção preventiva ou penas prolongadas.

Virginia Eubanks, em “Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor” (2018), sustenta que tecnologias automatizadas tendem a cristalizar desigualdades históricas, em vez de corrigi-las. Por meio de estudos de caso em Indiana, Los Angeles e Pittsburgh, ela demonstra como sistemas digitais reforçaram padrões discriminatórios contra populações vulneráveis (EUBANKS, 2018). Em sintonia, a Recomendação sobre Ética da Inteligência Artificial da UNESCO (2021) afirma que todos os atores do ciclo de vida da IA devem promover diversidade e inclusão, exigindo medidas proativas para evitar discriminação (UNESCO, 2021; MONTEIRO, 2021).

No Brasil, episódios de discriminação algorítmica também vêm à tona. Ferramentas de reconhecimento facial na segurança pública têm exibido enviesamentos raciais: o sistema adotado pela Polícia Civil do Rio de Janeiro (2019-2022) apontou erroneamente uma mulher negra como foragida em Copacabana; descobriu-se depois que a verdadeira suspeita estava presa havia quatro anos, demonstrando tanto desatualização da base quanto viés algorítmico (LEMOS, 2020).

O setor financeiro nacional traz outro exemplo. Como observa LEMOS (2020), mecanismos de pontuação são frequentemente sigilosos: “embora algumas pontuações, como crédito, estejam disponíveis ao público, os bureaus se recusam a revelar o método e a lógica de seus sistemas preditivos” (LEMOS, 2020, p. 15). Tal opacidade impossibilita contestação e pode afetar desproporcionalmente grupos vulneráveis, históricamente. Nessas circunstâncias, o Direito cumpre papel crucial: além de aferir justiça e conformidade normativa, deve funcionar simultaneamente como freio e indutor de inovações que promovam o desenvolvimento coletivo (BRASIL, Ministério da Saúde, 2023; LEMOS, 2020).

A LGPD brasileira representa passo importante contra a discriminação por meio de algoritmos. O artigo 6º, inciso IX, consagra o princípio de não discriminar, quando proíbe tratar dados para fins que discriminem atos ilícitos ou abusivos; o artigo 20, §2º, confere direito revisão as decisões geradas de forma automatizada. Pesquisadores, porém, destacam limitações: o texto legal não define claramente o que seja “discriminação ilícita ou abusiva”, e o direito à revisão não implica, necessariamente, acesso à explicação do algoritmo, dificultando a contestação de práticas discriminatórias (BRASIL, Ministério da Saúde, 2023; MENDES e MATTIUZZO, 2021; MONTEIRO, 2021).

Portanto, enfrentar a discriminação algorítmica no Brasil requer superar entraves institucionais e sociais profundos. A histórica desigualdade do país, aliada à captura de instituições por interesses de terceiros, cria ambiente propenso à perpetuação de injustiças por sistemas automatizados.

3.3 Diluição da responsabilidade institucional

O terceiro grupo de riscos se encontra à dispersão, e consequente enfraquecimento, da accountability. JIANG e NAUMOV (2023) analisam de modo formal o conceito de responsibility gap e diffusion em cenários de decisão coletiva, mostrando que, quando humanos e máquinas dividem tarefas, a responsabilidade acaba compartilhada a tal ponto que, na prática, nenhuma pessoa ou instituição é claramente apontada por falhas, erros ou danos provocados pelas decisões algorítmicas. Nos sistemas que atualmente utilizam IA, sobretudo aqueles baseados em AM, a cadeia causal que liga entrada, processamento e saída tende a ser opaca. A ideia de diffusion of responsibility descreve como a incorporação de algoritmos turva essa linha causal, pulverizando obrigações morais e jurídicas e tornando complexo identificar um único responsável ou adotar medidas corretivas eficazes. Esse cenário favorece o chamado “círculo do blame” – ou bystander effect – em que a culpa se dilui entre vários atores (BUSUIOC, 2020). Como resultado, surge um accountability gap: nenhum agente detém conhecimento e controle suficientes para ser plenamente responsabilizado pelas decisões do sistema, situação especialmente crítica no serviço público, onde a responsabilização é pilar democrático essencial (BUSUIOC, 2020).

A experiência da Estônia com sistemas de predição educacional[2] ilustra bem o problema. Reconhecido por sua vanguarda em governo digital, o país aplica IA em

[2] https://e-estonia.com/estonia-and-automated-decision-making-challenges-for-public-administration/

vários setores, inclusive na educação. O sistema nacional e-kool[3] emprega algoritmos para prever desempenho escolar e risco de evasão. Quando, porém, as previsões incorretas afetam a trajetória de um estudante, surge a pergunta: quem responde – o programador que escreveu o código, o gestor que autorizou a adoção, o diretor que seguiu as recomendações ou a própria instituição estatal que mantém o sistema? Para enfrentar essa complexidade, o governo estoniano editou, em 2021, um marco regulatório[4] específico para IA no setor público, estabelecendo responsáveis identificáveis e procedimentos de revisão de decisões automatizadas. O caso estoniano mostra, ao mesmo tempo, a dimensão do desafio e caminhos possíveis para fechar o accountability gap.

No Brasil, a difusão de responsabilidade coloca dificuldades particulares ao arcabouço jurídico. A Lei 9.784/99, que rege o processo administrativo federal, fixa princípios de responsabilização individual e institucional. O artigo 2º determina que a Administração deve observar, entre outros, motivação e indicação dos pressupostos de fato e de direito que sustentam a decisão, além do contraditório (BRASIL, 1999). De forma paralela, o artigo 37 da Constituição Federal consagra legalidade, impessoalidade, moralidade, publicidade e eficiência, pressupostos que exigem identificação nítida dos autores das decisões administrativas e de seus fundamentos (BRASIL, 1988). A delegação de escolhas a sistemas automatizados desafia, portanto, esses pressupostos tradicionais e, até o momento desta pesquisa, não há jurisprudência consolidada no país acerca da responsabilidade por danos produzidos por algoritmos estatais.

3.4 Diretrizes para uma governança algorítmica responsável

Construir uma governança algorítmica compatível em relação ao princípios democráticos demanda abordagem sistêmica, que vá além de soluções meramente técnicas. À luz dos riscos mapeados e das experiências nacionais e estrangeiras, esta seção apresenta um framework que se apoia em três pilares centrais, cada qual destinado a enfrentar dimensões específicas do desafio governamental, com foco especial nas realidades e exigências do Poder Executivo brasileiro.

3.4.1 Explicabilidade e transparência

O primeiro pilar da governança algorítmica coerente com valores democráticos é assegurar explicabilidade e transparência reais nos sistemas de decisão automática. Isso requer arranjos institucionais que permitam a qualquer interessado compreender em termos acessíveis, por que um algoritmo chegou a determinado resultado, um requisito especialmente relevante no sistema jurídico brasileiro.

A garantia de obtenção de informações claras está solidamente amparada tanto pela Constituição quanto pela legislação infraconstitucional. O art. 37 da Constituição

[3] https://e-estonia.com/what-is-e-education/
[4] https://www.kratid.ee/en/_files/ugd/980182_4434a890f1e64c66b1190b0bd2665dc2.pdf

Federal consagra a publicidade como um dos fundamentos da gestão pública; por sua vez, a Lei 9.784/99 exige que as decisões administrativas sejam devidamente fundamentadas. O art. 50º dessa mesma lei prevê que “os atos administrativos deverão ser motivados, com indicação dos fatos e dos fundamentos jurídicos”, abrindo espaço para requerer justificativas também de decisões baseadas em algoritmos. A Lei Geral de Proteção de Dados (LGPD) aprofunda essa lógica: o artigo 20º assegura ao titular o direito de revisar decisões automatizadas, enquanto o artigo 9º lhe garante “acesso facilitado”, em linguagem clara, às informações sobre o tratamento de seus dados. Ainda que a palavra “explicação” não apareça literalmente na LGPD, o conjunto desses dispositivos cria um ambiente propício à edificação de uma accountability transparente (BRASIL, Ministério da Saúde, 2023; MONTEIRO, 2021).A Estratégia Brasileira de Inteligência Artificial (EBIA), instituída pela Portaria MCTI nº 4.617/2021, eleva a explicabilidade a princípio essencial para o uso de IA no setor público, afirmando:

> Aspecto fundamental desse processo é estabelecer mecanismos que permitam prevenir e eliminar os vieses, que podem decorrer tanto dos próprios algoritmos utilizados, como também das bases de dados usadas para o seu treinamento. Para que um algoritmo seja ‘explicável’ ou ‘interpretável’, é desejável que as etapas do processo de aprendizado de máquina que resultaram em uma inferência sejam rastreáveis e que as variáveis que pesaram na tomada de decisão possam passar por escrutínio (MCTI, 2021, p. 24).

O mesmo documento orienta a adoção do princípio da precaução, recomendando avaliações específicas para aplicações classificadas como de alto risco por exemplo, sistemas utilizados em procedimentos que envolvem risco à vida no setor de saúde ou em monitoramento de espaços públicos para fins de segurança (BRASIL, Ministério da Saúde, 2023; MCTI, 2021).

Nesse contexto, a auditoria algorítmica desponta como instrumento indispensável à transparência. Conforme guia do Banco Interamericano de Desenvolvimento (BID), trata-se da avaliação completa de um Sistema de Decisão Automatizada (ADS) em todas as etapas de seu ciclo de vida: formulação, escolha de dados de treinamento, lógica decisória e resultados. O processo obedece a princípios como precisão, justiça, mitigação de vieses, não discriminação, privacidade e segurança da informação, com o propósito central de verificar aderência a padrões éticos, normativos e técnicos (BID, 2022). Normalmente, envolve análise documental, entrevistas com a equipe técnica, mapeamento das partes interessadas e relatórios públicos que recomendam ajustes para garantir conformidade legal e social.

As auditorias, por sua finalidade, podem assumir diferentes formatos:

- Auditorias de desempenho: avaliam acurácia, eficiência e robustez técnica (BID, 2022).
- Auditorias de conformidade: checam aderência a marcos regulatórios como LGPD, GDPR ou normas setoriais (UK, 2020).
- Auditorias adversariais: buscam vulnerabilidades, riscos sistêmicos e falhas éticas latentes (ETICAS FOUNDATION, 2023).

Um obstáculo recorrente, porém, é a complexidade dos modelos de aprendizado de máquina, muitas vezes tratados como “caixas-pretas”, o que dificulta a tradução de seu raciocínio (BARREDO e ARRIETA et al., 2020). Ainda assim, experiências concretas mostram avanços. No Brasil, o ALICE (Analisador de Licitações e Editais, CGU/TCU) foi submetido a avaliações que visaram fortalecer sua governança e aumentar a capacidade de detectar vieses e riscos em compras públicas (MACHADO, 2025). Conforme descreve a autora:

> ...às aquisições consideradas de maior criticidade e risco, definidas por critérios preestabelecidos de materialidade, criticidade e relevância, a ferramenta Alice deflagra, de maneira automática, procedimentos de auditoria denominados Avaliações Preventivas de Contratações. Estas ações são implementadas com o propósito de que os auditores da Controladoria-Geral da União (CGU) possam confirmar ou refutar os alertas previamente identificados pelo sistema. (MACHADO, 2025, p. 64)

O ALICE, portanto, desempenha papel crucial ao possibilitar identificação precoce de fraudes, irregularidades ou falhas que comprometam licitações. A abertura automática das Avaliações Preventivas de Contratações sinaliza postura proativa da CGU, voltada a evitar prejuízos ao erário e a aprimorar a gestão administrativa (MACHADO, 2025).

3.4.2 Supervisão humana qualificada

Coadunando com a seção anterior, o segundo pilar da governança algorítmica destaca a indispensabilidade de uma supervisão humana devidamente qualificada em cada etapa do processo de tomada de decisão por sistemas automatizados. Essa supervisão vai para além da aprovação meramente formal, exigindo competências especializadas e arranjos institucionais apropriados ao contexto do Poder Executivo federal. No caso do ALICE (Analisador de Licitações e Editais) da CGU, já mencionado, vigora um modelo de cooperação no qual algoritmos identificam possíveis irregularidades e auditores humanos realizam investigações aprofundadas. A divisão de tarefas permite que a máquina trate grandes volumes de contratos enquanto os especialistas concentram seu conhecimento nos casos sinalizados como problemáticos (MONTEIRO, 2021).

A literatura especializada recomenda amplamente modelos híbridos, nos quais a inteligência artificial serve de suporte sem substituir o discernimento humano, pois eles combinam eficiência técnica com controle social responsável (BARREDO e ARRIETA et al., 2020; CENTRE FOR DATA ETHICS AND INNOVATION, 2020). A interação entre análise estatística algorítmica e revisão humana é crucial para evitar erros sistêmicos, vieses e o chamado automation bias[5] – tendência dos operadores a

[5] https://cset.georgetown.edu/publication/ai-safety-and-automation-bias/#:~:text=Automation%20bias%20is%20the%20tendency,the%20face%20of%20contradictory%20information.

confiar exageradamente nos sistemas automáticos, negligenciando sua própria avaliação.

Tal arranjo aproxima-se da “transparência temperada” proposta por Cunha, Stamile e Grupenmacher (2024). Essa forma de transparência corresponde à “transparência qualificada” de Pasquale (2017), uma via intermediária entre sigilo algorítmico absoluto e divulgação total – extremos que podem ser inadequados dependendo do caso. Ela se concretiza por meio de análises e certificações, realizadas por autoridades competentes, acerca da segurança, qualidade, validade e confiabilidade do sistema (Pasquale, 2017). No Brasil, esse procedimento se conecta à possibilidade de auditoria pela ANPD, prevista no § 2º do artigo 20 da LGPD (CUNHA, STAMILE e GRUPENMACHER, 2024, p. 11).

De modo análogo, o SISAM não executa sozinho as atividades de fiscalização; ele oferece subsídios que são avaliados e, se pertinente, ratificados pelos auditores fiscais. Essa arquitetura preserva o juízo crítico dos profissionais, sobretudo em cenários sensíveis em que os dados históricos não refletem todas as particularidades do caso concreto (CUNHA; STAMILE; GRUPENMACHER, 2024; RECEITA FEDERAL DO BRASIL, 2023). Vale notar, contudo, que iniciativas desse tipo enfrentam obstáculos práticos, como a alta rotatividade em cargos de confiança e a disparidade de maturidade técnica entre órgãos públicos federais (CUNHA; STAMILE; GRUPENMACHER, 2024).

3.4.3 Responsabilização institucional e controle social

A literatura internacional dedicada à governança democrática da inteligência artificial, sobretudo na esfera pública, insiste na necessidade de dispor de mecanismos robustos de responsabilização institucional e de controle social para sistemas algorítmicos (BUSUIOC, 2020). No Brasil, tal exigência encontra respaldo no art. 37, § 6º, da Constituição Federal, que consagra a responsabilidade objetiva da Administração. Esse mesmo princípio é apontado em estudos estrangeiros como instrumento de reparação para a “lacuna de responsabilidade” (accountability gap) ocasionada pela opacidade técnica e pela fragmentação das decisões em cadeias de IA (BUSUIOC, 2020; BARREDO e ARRIETA et al., 2020).

Tornar efetiva a responsabilização por atos automatizados exige, porém, regulamentação infraconstitucional específica, apta a lidar com a causalidade difusa ou incerta típica de sistemas algorítmicos complexos (BUSUIOC, 2020; BARREDO e ARRIETA et al., 2020). As medidas reparatórias devem ir além da compensação financeira, incorporando correção ágil de registros, revisão administrativa das decisões geradas por IA e salvaguardas preventivas para evitar recorrências (BID, 2022). Nesse sentido, boas práticas internacionais recomendam criar portais públicos unificados que concentrem avaliações de impacto, relatórios de auditoria e canais formais para denúncias ou pedidos de correção relacionados ao uso de algoritmos estatais (BID, 2022).

Decisões com elevado grau de discricionariedade, forte impacto sobre direitos fundamentais ou dependência intensa de contexto não devem ser totalmente automatizadas; nesses casos, a supervisão e eventual revisão por profissionais qualificados permanece indispensável (BARREDO E ARRIETA et al., 2020; VEALE e BINNS, 2017). O AI Act europeu sugere mapeamento institucional para determinar,

setor a setor, quais deliberações comportam automação total, parcial ou devem manter-se estritamente humanas (EUROPEAN PARLIAMENT AND COUNCIL, 2024; LOI e SPIELKAMP, 2021). Já o modelo canadense de avaliação de impacto algorítmico enfatiza participação social, inclusão de grupos vulneráveis e revisões periódicas (Canada, 2019). Com vigência desde agosto de 2024, o AI Act impõe avaliação obrigatória de risco e transparência proporcionada ao nível de perigo do sistema, oferecendo referência regulatória compatível com as necessidades brasileiras (EUROPEAN PARLIAMENT AND COUNCIL, 2024).

Considerando a heterogeneidade das instituições nacionais, recomenda-se percurso evolutivo: iniciar pelos setores mais maduros no uso de IA e, gradualmente, expandir práticas de responsabilização e de participação pública aos demais domínios governamentais (LOI e SPIELKAMP, 2021; BANCO INTERAMERICANO DE DESENVOLVIMENTO, 2022).

4. Considerações Finais

A implementação de ferramentas de inteligência artificial (IA) pelo poder público transfere parcela significativa da deliberação estatal para modelos estatísticos de funcionamento pouco visível. Embora apresentem promessas de ganho de eficiência, tais sistemas ainda se mostram ambivalentes diante das exigências do Estado Democrático de Direito. A discussão desenvolvida neste artigo demonstrou, primeiro, que a opacidade algorítmica não desaparece mesmo quando o software é desenvolvido internamente. Tanto o SISAM, da Receita Federal, como o ALICE, da Controladoria-Geral da União, evidencia que a lógica de aprendizado de máquina permanece inacessível ao cidadão comum e, em diversos casos, aos próprios gestores responsáveis. Essa invisibilidade conflita com o princípio constitucional da publicidade e com o dever de motivação dos atos administrativos, que impõe a explicitação dos fatos e fundamentos de cada decisão. Além disso, experiências internacionais revelam que a abertura integral do código-fonte pode produzir efeitos indesejáveis, motivo pelo qual se consolida a ideia de transparência qualificada: relatórios públicos, registros detalhados de uso, auditorias independentes e divulgação criteriosa de parâmetros sensíveis formam o patamar mínimo para recuperar a inteligibilidade dos sistemas.

Em segundo lugar, vieram à tona os riscos de vieses estruturais e discriminação automatizada. Pesquisas sobre o COMPAS, nos tribunais norte-americanos, e sobre ferramentas de reconhecimento facial empregadas por forças policiais brasileiras mostram que modelos estatísticos tendem a absorver e amplificar desigualdades históricas, colocando grupos vulneráveis em situação de maior exposição a erros. A Lei Geral de Proteção de Dados oferece salvaguardas relevantes, como o direito à revisão de decisões automatizadas e o princípio da não discriminação, mas carece de parâmetros operacionais que indiquem quando a discriminação passa a ser ilícita ou abusiva. Daí a necessidade de avaliações de impacto algorítmico, testes sistemáticos de viés nas bases de dados e monitoramento humano permanente como complementos indispensáveis ao arcabouço jurídico.

O terceiro achado refere-se ao chamado accountability gap, lacuna de responsabilização decorrente da pulverização de responsabilidades entre desenvolvedores, fornecedores, gestores públicos e usuários finais. Quando um

sistema automatizado causa prejuízo, a cadeia decisória fragmentada dificulta a identificação de culpados e a reparação das vítimas. A experiência da Estônia, que já dispõe de regras específicas para IA no setor público, demonstra ser possível atribuir responsáveis formais em cada etapa do ciclo de vida do algoritmo e submeter soluções de alto risco a análises prévias e planos de contingência. No Brasil, a responsabilidade objetiva prevista no art. 37, § 6º, da Constituição serve de alicerce, mas a falta de regulamentação infraconstitucional voltada ao contexto algorítmico impede respostas céleres e proporcionais.

Esses resultados convergem para um desenho de governança algorítmica estruturado em três pilares. O primeiro exige níveis suficientes de explicabilidade e transparência que permitam a cidadãos, órgãos de controle e tribunais compreender, ainda que parcialmente, a lógica e os efeitos dos algoritmos em uso. O segundo impõe supervisão humana qualificada desde a concepção até a operação cotidiana, de modo a mitigar vieses e resguardar o juízo de proporcionalidade. O terceiro reforça a responsabilização institucional, por meio de portais de prestação de contas, ampliação do direito de contestação e sanções compatíveis com os danos causados. Entre as medidas práticas sugeridas estão: (1) tornar obrigatória a auditoria de desempenho, conformidade e risco antes da implantação de sistemas de IA de alto impacto; (2) exigir cláusulas contratuais de “transparentização” na aquisição de softwares privados; e (3) criar um cadastro nacional unificado que permita ao cidadão verificar quais sistemas estão ativos, com que finalidade e sob quais salvaguardas operam.

A principal limitação da pesquisa reside no uso predominante de fontes secundárias, o que impediu captar a perspectiva de desenvolvedores privados, usuários diretamente afetados e possíveis atualizações recentes do SISAM e do ALICE. Investigações futuras devem recorrer a metodologias mistas, combinando etnografia institucional, experimentos sobre percepção de justiça e métricas quantitativas de viés, para desvelar camadas de desigualdade ainda invisíveis nos relatórios técnicos. Mesmo assim, os achados reforçam um ponto crucial: a governança de algoritmos deixará de ser privilégio dos especialistas apenas quando se apoiar em explicabilidade técnica, supervisão humana contínua e responsabilização efetiva, condições essenciais para harmonizar inovação digital e valores democráticos.