

MACHINE LEARNING, INOVAÇÕES GOVERNAMENTAIS E NOVOS DESAFIOS PARA A TRANSPARÊNCIA PÚBLICA

MACHINE LEARNING, GOVERNMENT INNOVATIONS AND NEW CHALLENGES FOR PUBLIC TRANSPARENCY

Fernando Filgueiras 1

Resumo: Governos têm produzido inovações importantes em rotinas administrativas e no processo das políticas públicas com a crescente adoção de sistemas baseados em algoritmos de machine learning. Movidos por uma ideia ampla de transformação digital, a adoção de sistemas de machine learning para tomar decisão e realizar tarefas da administração pública criam uma nova camada de complexidade para as políticas de transparência. Algoritmos de machine learning desafiam as políticas de transparência à medida que o desenho de sistemas e o modo como os algoritmos calculam cursos de ação pública são opacos para a sociedade. Algoritmos de machine learning implementados em diferentes atividades governamentais implicam em desafios para a transparência e accountability. Este artigo discute estes desafios e aponta caminhos com relação à promoção da transparência. Mais do que se concentrar no processo de desenho de arquiteturas algorítmicas, policymakers devem se concentrar também na transparência dos resultados de sistemas aplicados em governos. Processos de autorregulação de IA são insuficientes para enfrentar os desafios conexos do avanço e crescente adoção de algoritmos de machine learning em inovações governamentais.

Palavras-chave: Algoritmos. Machine Learning. Transparência. Accountability. Transformação Digital.

Abstract: Governments have produced essential innovations in administrative routines and the public policy process with the increasing adoption of systems based on machine learning algorithms. Driven by a broad idea of digital transformation, adopting machine learning systems to make decisions and carry out public administration tasks creates a new layer of complexity for transparency policies. Machine learning algorithms challenge transparency policies as systems design and how algorithms compute public courses of action are opaque to society. Furthermore, machine learning algorithms implemented in different government activities imply challenges for transparency and accountability. This article discusses these challenges and points out ways to promote transparency. More than designing algorithmic architectures, policymakers should also focus on the transparency of the results of systems applied in governments. AI self-regulation processes are insufficient to address the related challenges of advancing and increasing the adoption of machine learning algorithms in government innovations.

Keywords: Algorithms. Machine Learning. Transparency. Accountability. Digital Transformation.

1 Doutor em Ciência Política pelo Instituto Universitário de Pesquisa do Rio de Janeiro (IUPERJ). Professor associado da Universidade Federal de Goiás (UFG). Affiliate faculty da Indiana University. Pesquisador do Instituto Nacional de Ciência e Tecnologia em Democracia Digital. Lattes: <http://lattes.cnpq.br/4348768274838297>. ORCID: <https://orcid.org/0000-0001-9570-8113>. E-mail: fernandofilgueiras@ufg.br

Introdução

Tecnologias digitais, especialmente a inteligência artificial, têm promovido disrupções na forma como governos formulam, implementam e avaliam políticas públicas. Reformas na administração pública e nas políticas públicas são introduzidas silenciosamente por meio da adoção gradativa de sistemas algorítmicos (Filgueiras, 2023). De um lado, a administração pública concretiza uma série de mudanças com base em um processo crescente de plataformização do governo. De outro lado, nas políticas públicas, temos uma mudança epistêmica na forma como decisões são tomadas e rotinas organizacionais são refeitas com o apoio de sistemas sociotécnicos baseados em algoritmos de *machine learning*. Algoritmos de *machine learning* ativaram um imaginário sociotécnico em governos, transformando e racionalizando diversos aspectos organizacionais de funcionamento da máquina administrativa.

Algoritmos de *machine learning* gradativamente se tornam pervasivos em diferentes dinâmicas de governo, alterando vários aspectos da governança do setor público. Algoritmos de *machine learning* estão no centro dessas transformações, provocando mudanças disruptivas, as quais podem ser compreendidas como alterações bruscas no modo de fazer e compreender problemas de políticas e construir soluções (Jones e Baumgartner, 2005). Empacotados em uma ideia ampla de transformação digital, muitos aspectos da governança pública são transformados, fazendo com que algoritmos de *machine learning* operem no coração do processo decisório analisando dados e construindo informações aplicadas em diversos aspectos das políticas públicas.

O conjunto dessas mudanças é bastante amplo e provoca incertezas e ambiguidades quanto ao seu real alcance e capacidade de promover melhorias efetivas na vida das pessoas, representando impactos sociais diversos. A pervasividade de sistemas algorítmicos em governos cresce à medida do avanço de ideias e políticas de transformação digital, assim como no maior volume e velocidade com que dados são coletados, armazenados e processados em escala. Essa pervasividade se dá no modo como algoritmos de *machine learning* cada vez mais criam as condições de informações do processo decisório em diferentes domínios de políticas públicas.

Por exemplo, as decisões de políticas do Programa RenovaBio, do Ministério de Minas e Energia, são algorítmicamente moldadas. O Programa RenovaBio, instituído em 2017, é uma política de natureza regulatória, cujo principal instrumento é a certificação de produtores e importadores de biocombustíveis. O objetivo do programa é reduzir a emissão de carbono na atmosfera, contribuindo para políticas climáticas. O programa leva em consideração a relação entre eficiência energética e redução de gases de efeito estufa com o objetivo de descarbonizar a matriz energética de transportes, além da previsibilidade requerida pelo mercado. A certificação é um instrumento importante da política à medida que incentiva o uso de créditos de carbono e negociação de ações em bolsas de valores. A certificação gira em torno de notas de eficiência energético-ambiental calculadas por um algoritmo de *machine learning*. Ou seja, a decisão pela certificação ou não de produtores, assim como as consequências dessa decisão, são algorítmicamente moldadas.

Decisões tomadas por algoritmos de *machine learning* não se restringem a questões de mercado. Eles também se tornam presentes na área de políticas sociais. O Instituto Nacional de Seguridade Social (INSS) centraliza as políticas de proteção social com especial enfoque no sistema de aposentadorias e benefícios sociais. Como resposta para o aumento de demanda da sociedade para concessão e revisão de benefícios sociais, o INSS passou a adotar algoritmos de *machine learning* que analisem os requerimentos. Dizendo em outras palavras, um algoritmo de *machine learning* analisa o requerimento textual e os dados de registros administrativos do INSS para conceder ou não o benefício. Em estimativa do Tribunal de Contas da União, cerca de 60% dos pedidos são recusados, criando, na expressão do Ministro Aroldo Cedraz, um sistema de “desproteção social” no Brasil. Segundo o Ministro, em decisão exarada em acórdão do TCU, “É essencial que algoritmos sejam validados de forma criteriosa, com seus resultados sendo submetidos à revisão humana pelo período necessário para se obter segurança de que não existem falhas detectáveis que possam trazer prejuízo a qualquer das partes” (Tribunal de Contas da União, 2023). De forma geral, algoritmos transformam as ideias por trás das políticas de proteção social. O uso extensivo de algoritmos de *machine learning* na proteção social cria uma perspectiva liberal restrita de concessão de benefícios e uma concepção ambígua atrelada à desconfiança. Ou seja,

uma concepção que se concentra na auditoria dos benefícios e desmantelamento da efetividade da proteção, em uma tendência mundial de desmantelamento do estado de bem-estar por meio da digitalização (Collington, 2022).

Ambos os casos ilustram como algoritmos de *machine learning* vão se tornando pervasivos no processo decisório das políticas públicas e em rotinas organizacionais da administração pública. Nesse contexto, algoritmos de *machine learning* funcionam como instrumentos de racionalização que modificam a forma como as coisas são feitas e criam novos enquadramentos para ideias no setor público. A administração pública e as políticas públicas gradativamente se tornam programáveis por meio de algoritmos, criando, por sua vez, novos dilemas sociais (Filgueiras, 2022). Ambientes organizacionais ditos inteligentes, que emergem com os algoritmos de *machine learning*, criam novas situações de ação, em que outputs desses sistemas “programam” a ação humana com implicações coletivas diversas (Mendonça; Filgueiras; Almeida, 2023; Frischmann; Selinger, 2018).

No contexto de toda essa transformação que emerge com a adoção de algoritmos de *machine learning* no processo das políticas públicas e da administração pública, um novo problema emerge. Embora tenha havido avanços desde o início dos anos 2000 na construção da transparência pública e fortalecimento da *accountability*, a adoção de algoritmos de *machine learning* torna políticas e práticas administrativas mais opacas, fazendo com que cidadãos sejam incapazes de examinar e entender o processo decisório. Algoritmos de *machine learning* são “caixas pretas” que atuam diretamente no processo decisório sem que cidadãos possam entender como as decisões foram tomadas (Pasquale, 2015).

O objetivo desse artigo é discutir como inovações governamentais que emergem com a adoção de algoritmos de *machine learning* criam novas camadas e dilemas para o fortalecimento de políticas de transparência. Ou seja, algoritmos de *machine learning* impactam diretamente a transparência, tornando as políticas públicas e práticas administrativas mais opacas e difíceis de controlar. Partindo dessa premissa, o artigo trata dos seguintes pontos: Primeiro, o que são algoritmos de *machine learning* e como eles transformam rotinas organizacionais do Estado? Segundo, quais são os desafios democráticos para a transparência que emergem com a crescente adoção de algoritmos de *machine learning*? Terceiro, como regular estes algoritmos e quais as discussões emergentes sobre esse tópico. O artigo conclui essa discussão tratando dos avanços institucionais que surgem com novas práticas que precisam ser observadas no processo de transformação digital de governos.

O que são e quais as implicações de algoritmos de machine learning?

Um algoritmo pode ser descrito como sequências finitas de instruções ou etapas bem definidas para resolver um determinado problema, tarefa ou cálculo (Knuth, 1968). Eles representam regras de cálculo que definem procedimentos formais a partir dos quais uma entrada de dados produz uma saída na forma de informação. Os avanços que têm sido produzidos com a reconstrução da inteligência artificial têm colocado os algoritmos no centro dos debates públicos a respeito de sistemas sociotécnicos. Algoritmos de *machine learning*, em específico, representam um método de inteligência artificial.

Inteligência artificial (IA) não é propriamente uma coisa e nem há consenso no campo sobre seu conceito e prática (Wang, 2019). A IA não possui um conceito preciso na literatura especializada. As bases conceituais da IA estão na inteligência para realizar uma tarefa ou resolver um problema, que pode variar de acordo com o conceito de inteligência que está sendo empregado (Russell, 2019). O pressuposto do desenvolvimento da IA é a imitação (Turing, 1950). Se a inteligência representa a racionalidade humana, a IA é uma solução que mimetiza essa racionalidade (Simon, 1957). Se a inteligência for concebida como uma função cognitiva, a IA é a solução que busca a expansão do conhecimento e a produção de aprendizagem (Russell; Norvig, 2010). Se inteligência é entendida como capacidade, os sistemas de IA criam capacidades para realizar tarefas complexas (Minsky, 1985). Se inteligência é pensamento estruturado, as soluções de IA buscam imitar essa estrutura para resolver diferentes problemas mais rapidamente (Markram, 2006).

IA, portanto, pode significar coisas diferentes dependendo da metodologia que for empregada

para resolver um problema. O conceito estabelecido para IA é o da capacidade de uma máquina de aprender com dados massivos - coletados, armazenados e processados com metodologias de big data - para prever e simular diferentes respostas. IA significa a constituição de agentes que percebem e agem para atingir um objetivo, adotando diversas metodologias e abordagens. Esses agentes percebem e constroem um fluxo de entradas que é convertido em um fluxo de ações - respostas (Russell, 2019). Inteligência artificial diz respeito a um campo de conhecimento, por meio do qual máquinas otimizam a obtenção de informações para otimizar, por sua vez, a tomada de decisão e a realização de tarefas (Simon, 1957, 1995).

Uma forma de entender a imitação da inteligência humana é construindo uma analogia com uma situação prática. Imagine, leitor, que você está em casa pronto para sair para o trabalho. Quando você abre a porta de sua casa e olha para o ambiente, percebe um conjunto de nuvens escuras e ouve uma série de barulhos que você classifica como trovões. Ao longo de sua vida, enumerando exemplos de como você se molhou em uma situação como essa, ou simplesmente porque seus pais te ensinaram que nuvens escuras e trovões aumentam a probabilidade de chuva, você terá uma razão suficiente para retornar à sua casa e pegar um guarda-chuva. Você aprendeu ao longo de sua vida pessoal e calculou instintivamente a probabilidade de chuva e com essa informação construiu uma ação prática de pegar o guarda-chuva. Ou seja, se eu percebo, calculo rotas e conheço que vai chover, pego um guarda-chuva. O mesmo pressuposto de agência está embutido nos algoritmos de *machine learning*.

No pressuposto da imitação da inteligência humana, algoritmos de *machine learning* fazem o mesmo que essa situação descrita acima. O algoritmo representa um “book of rules” para estabelecer procedimentos formais para calcular um determinado curso de ação com os dados de entrada (Turing, 1950). Existem diferentes tipos de algoritmos para diferentes tipos de aplicações. Os algoritmos de *machine learning* trabalham com a ideia de que a máquina pode aprender com os dados e produzir outputs informacionais que moldam a agência (Samuel, 1959). O estudo e o uso de algoritmos são multidisciplinares e têm despertado o interesse de pesquisadores de praticamente todas as áreas da ciência, desde a ciência da computação, matemática e engenharia até as ciências sociais. Os cientistas da computação e engenheiros tendem a focar seu trabalho no desenho e análise de algoritmos, determinando sua eficiência, correção e otimização de uma perspectiva técnica.

A emergência de big data possibilitou reconstruir a área de inteligência artificial e otimizar a presença de algoritmos de *machine learning* para resolver diferentes problemas. Em vez de especificar regras e rotas, passo a passo que os algoritmos tradicionais seguem, os algoritmos de *machine learning* são softwares que aprendem com exemplos, dados e experiências (Samuel, 1959). Estes algoritmos calculam funções de aprendizado e são capazes de aprender com os dados para resolver problemas. Os algoritmos de *machine learning* fazem parte de uma família de métodos de inteligência artificial (IA) que incluem aprendizado profundo e aprendizado por reforço (Kelleher, 2019). Estes algoritmos estão embarcados em diferentes sistemas implementados em assistentes virtuais, veículos autônomos e sistemas de decisão.

Essencialmente, um algoritmo de *machine learning* recebe dados como entrada e produz um modelo que representa os padrões que o algoritmo aprendeu com esses dados (Samuel, 1962). Com base na estatística e na teoria da probabilidade, os algoritmos de *machine learning* geram resultados com algum grau de incerteza sobre o que é útil ou ideal, criando condições de agência para realizar diferentes tarefas. Algoritmos de *machine learning*, também conhecidos como algoritmos de autoaprendizagem, definem um processo para analisar um conjunto de dados e identificar padrões de dados recorrentes.

Implementados em diferentes interfaces ou plataformas, algoritmos de *machine learning* proporcionam, além do aprendizado estático com os dados, perspectivas dinâmicas baseadas na coleta de *feedbacks*. Estes *feedbacks* aprimoram o aprendizado por reforço e, por conseguinte, a acurácia de sistemas algorítmicos para lidar com questões transacionais e interativas de humanos que estejam conectados online (Kelleher, 2019). Em outras palavras, algoritmos de *machine learning* aplicados em diversas políticas públicas proporcionam meios técnicos para que formuladores colem *feedbacks* em tempo real e, assim, aprimorem e mudem efeitos de implementação de políticas por meio de sistemas algorítmicos.

O aprendizado por reforço concentra-se em aprender com a experiência e situa-se entre o aprendizado não supervisionado e o supervisionado, fazendo com que a estrutura do algoritmo trabalhe com *feedbacks* proporcionados na interação entre humanos e máquinas (Kelleher, 2019). Algoritmos de *deep learning* – um subcampo de algoritmos de *machine learning* – são redes neurais artificiais inspiradas em insights extraídos da fisiologia e da neurociência. Depois que o modelo é treinado, a segunda parte dos algoritmos de aprendizado de máquina é a inferência, que se refere ao processo de usar o modelo para gerar novos valores para os exemplos (Kelleher, 2019).

De acordo com Issar e Aneesh (2022), o uso de sistemas algorítmicos no processo de governança das políticas públicas cria novas condições de agência, uma vez que sistemas sociotécnicos realizam o trabalho de decisão e implementação por meio de dados massivos processados por algoritmos de *machine learning*. Eles transformam as condições de agência porque a informação criada a partir dos dados tem implicações diretas na forma como humanos vão perceber, interagir e racionalizar situações práticas. Algoritmos de *machine learning*, calculando diferentes funções de aprendizado em diferentes situações, produzem mudanças epistêmicas na sociedade, transformando o conhecimento humano e suas formas de lidar com situações práticas (Coeckelbergh, 2022). Aplicados na governança pública, algoritmos de *machine learning* institucionalizam novas ordens políticas baseadas no cálculo da ação prática em contextos coletivos (Mendonça; Filgueiras; Almeida, 2023; Amooore, 2022).

Algoritmos de recomendação, algoritmos de identificação de imagens, mecanismos de busca, algoritmos de roteamento e “agentes” digitais são exemplos de algoritmos que limitam ou eliminam a necessidade de envolvimento humano em algumas tarefas. Algoritmos são implantados em instalações de fronteira, por exemplo, para substituir decisões humanas em operações de segurança (Amooore, 2021). Da mesma forma, algoritmos de *credit score* são usados para avaliar a adequação de um solicitante de empréstimo ou outras operações comerciais e financeiras e, também, para controle social pelo Estado (Dai, 2020).

Algoritmos de *machine learning* são desenhados hoje em escala industrial. A escolha da arquitetura de um algoritmo não é feita do zero e nem buscando uma solução correta para um determinado problema. Algoritmos de *machine learning* são soluções prontas buscando por problemas (Filgueiras, 2022). O desenho de um sistema algorítmico é feito por tentativa e erro buscando uma solução mais acurada na resposta dos dados. Retornemos ao exemplo ilustrativo do algoritmo do INSS que determina ou não o pagamento de benefícios sociais e aposentadorias. O seu desenho não buscou a solução mais correta com base nos fundamentos do Direito Previdenciário e Administrativo, mas a solução mais acurada para detectar fraudes. Ou seja, um algoritmo que retém processos de aposentadorias ou benefícios sociais de cidadãos é acurado para resolver os problemas de fraudes, conforme as entradas de dados disponibilizadas para ele. Porém, é incorreto como elemento de análise com respeito a direitos. Provavelmente, este algoritmo foi escolhido dentre diferentes algoritmos testados para detectar fraudes. O teste e validação do algoritmo não são a busca da solução correta, mas aquela que é estatisticamente mais acurada para responder ao problema, criando novos problemas.

A construção de sistemas algorítmicos para tomar decisões e realizar tarefas diversas no setor público lida com esse dilema. Embora possam representar uma solução acurada para resolver um problema, quando são aplicados na governança pública, os algoritmos de *machine learning* criam situações em que a escolha da arquitetura do sistema por desenvolvedores, bem como dos dados que serão utilizados e de como o algoritmo vai calcular as funções de aprendizado são totalmente opacas ao controle público. Essa opacidade não deriva apenas da escolha dos desenvolvedores. Ela implica na própria dinâmica da tecnologia, que é flexível o suficiente para mudar as funções de aprendizado por meio de *feedbacks* ou novos dados coletados. Ou seja, algoritmos de *machine learning* são flexíveis o suficiente para exacerbar problemas de transparência. A opacidade dos algoritmos obscurece a natureza política por trás das decisões públicas. Cidadãos não são capazes e nem são informados sobre como os desenvolvedores escolhem a arquitetura de algoritmos e de dados que impactam suas vidas (Gorwa; Binns; Katzenbach, 2020). Ordens políticas baseadas em *machine learning* são necessariamente opacas e nada *accountables* (Amooore, 2022).

Esta opacidade dos algoritmos de *machine learning* obscurece ainda questões relativas à equidade e justiça das políticas públicas. Bases de dados podem ser enviesadas de maneira que

prejudiquem pessoas mais vulneráveis na sociedade. Diferentes experiências são reportadas na forma como algoritmos de machine learning produzem asilos digitais que punem os mais pobres (Eubanks, 2018), produzem vieses raciais diversos que discriminam pessoas pretas (Benjamin, 2019; Obermayer *et al.*, 2019), são seletivos com relação à entrada de pessoas refugiadas de origem muçulmana em fronteiras (Amoore, 2021). Em diferentes situações, a opacidade dos algoritmos de *machine learning* retiram a agência epistêmica da sociedade, transferindo-a para sistemas sociotécnicos que moldam novas formas de interação social e produzem consequências políticas.

Ordens políticas construídas a partir de algoritmos de *machine learning* amplificam as incertezas públicas, criando novas demandas de transparência (Innerarity, 2021). As incertezas são amplificadas em democracias quando cidadãos não conseguem controlar o fluxo da decisão pública, criando situações epistêmicas que envolvem desconhecimento ou desinformação (Coeckelbergh, 2022). Precisamos lembrar que algoritmos de *machine learning* calculam funções de aprendizado por meio de modelos estatísticos, a partir das quais resolvem um problema e realizam um objetivo. A busca da melhor acurácia de resposta, do ponto de vista da escolha da arquitetura de sistemas, resulta na própria incerteza quanto aos resultados. Por outro lado, ordens políticas baseadas em algoritmos de *machine learning* produzem respostas ambíguas em termos de tarefas organizacionais da administração pública (Filgueiras, 2022). Ambiguidades se tornam presentes em inovações governamentais, amplificando situações de incerteza. Ou seja, sistemas sociotécnicos baseados em algoritmos de *machine learning*, desenvolvidos e aplicados para a execução de tarefas administrativas e processos decisórios de políticas em governos, são por definição opacos e produzem consequências diversas no funcionamento da democracia.

Machine learning implica em um desafio de democratização de ordens políticas que emergem a partir dos algoritmos como novas instituições da sociedade (Mendonça, Filgueiras e Almeida, 2023). Na próxima seção trataremos desse desafio e do modo como a transparência é vital no processo de desenvolvimento digital.

Machines of muddling through: O desafio democrático da accountability e transparência algorítmicas

Incertezas fazem parte do processo de formulação de políticas públicas. O processo da política pública não é um ciclo de racionalidade. Na definição do problema, na escolha de uma alternativa e na formulação da política, interferem valores e práticas políticas mais amplas que determinam o resultado do desenho institucional adotado (Lindblom, 1959). Além disso, o processo de produção da política pública não é linear e unidirecional, mas um processo incremental que envolve tentativas e erros sequenciais. De acordo com Lindblom, políticas públicas são formuladas em um processo incessante de tentativa e erro por meio de desenhos institucionais. *Policymakers* escolhem diferentes soluções em contextos de racionalidade limitada e sem uma teoria formal de programas, fazendo com que políticas públicas sejam uma ciência baseada em *muddling through*. Tentativa e erro é uma forma comum de decisão tanto com respeito ao desenho de uma política pública quanto do desenho de um algoritmo de *machine learning*. Nesse paralelo, podemos considerar que a adoção de um algoritmo de *machine learning* para resolver problemas de políticas carece dos mesmos problemas com relação à escolha de um desenho de política (Filgueiras, 2022). Estas escolhas são feitas em contextos de incertezas, podendo produzir respostas ambíguas para problemas públicos.

A escolha dos desenhos de sistemas é feita em contextos de racionalidade limitada, em que *policymakers* não necessariamente conseguem antecipar os resultados e produtos que emergem da aplicação de algoritmos de *machine learning* para resolver problemas. Existem incertezas, mas a adoção de algoritmos de *machine learning* é politicamente informada e moldada por ideias de transformação digital dirigidas pela redução dos custos da burocracia, respostas para situações de crise econômica e realização de princípios liberais. Entretanto, são estas escolhas de sistemas e o modo como eles operam que tornam opacas políticas e processos administrativos em governos digitais. Se antes a adoção de transparência visava a reduzir o déficit de informação entre cidadãos e gestores humanos, incluindo todos os dilemas de agência, a adoção de algoritmos de *machine*

learning em inovações governamentais cria uma camada nova de dificuldade para realizar o princípio democrático de *accountability* e transparência.

A *accountability* é um elemento crucial das democracias, se as democracias significam que o poder não deve ser concentrado em uma única instituição ou grupo de atores. As democracias exigem *accountability* no sentido de que uma instituição é responsabilizada por outras, devendo justificar as ações praticadas por meio dela e estar sujeita a eventuais formas de punição (Bovens, 2007; Mulgan, 2000). Esse é o cerne da ideia de freios e contrapesos, que permite o controle de uma instituição por outras e gera alguma forma de equilíbrio, evitando a concentração de poder. A transparência é condição necessária para a *accountability*, pois a fiscalização e o controle das instituições dependem da disponibilidade de informações sobre elas e suas ações (Filgueiras, 2016).

Um importante conjunto de teóricos democráticos enfatiza a crescente importância da *accountability* para as democracias contemporâneas. John Keane, por exemplo, fala do surgimento de uma *democracia monitorada* na segunda metade do século XX decorrente da crescente importância das redes formais e informais de vigilância sobre os atores políticos (Keane, 2013). Rosanvallon aponta a *contrademocracia* que historicamente contribuiu para o controle do poder. Os mecanismos de soberania negativa incluem não apenas vetos, mas também o monitoramento e vigilância por parte das instituições (Rosanvallon, 2008).

Algoritmos de *machine learning* criam novas ordens políticas e uma série de desafios para a promoção da transparência e da *accountability*. Em muitos aspectos, algoritmos de *machine learning* decidem sem que cidadãos tenham a possibilidade de supervisionar ou controlar sua ação. O uso de algoritmos de *machine learning* reforça ainda mais os dilemas de delegação que emergem da relação entre *principals* e agentes. Como agentes que percebem um fluxo de dados e constroem outputs de informação que mudam a agência epistêmica do serviço público, algoritmos de *machine learning* aprofundam as assimetrias informacionais. Cidadãos se vêem incapacitados de controlar a ação dos agentes, fazendo com que os interesses de organizações burocráticas se sobreponham aos interesses dos cidadãos (Eisenhardt, 1989). No caso da adoção de sistemas baseados em *machine learning*, temos múltiplos agentes, reforçando ainda mais os dilemas de transparência e *accountability*. Múltiplos agentes amplificam os dilemas de delegação (Gailmard, 2009).

A democratização dos algoritmos é fortemente dependente do estabelecimento de formas de *accountability* (Binns, 2018; Kroll et al, 2017). Essa democratização requer maneiras mais claras de responsabilizar os atores e torná-los responsáveis pelas consequências reais dos sistemas algorítmicos. Para isso, a transparência é sempre apontada como elemento chave. Se outras instituições e atores não puderem entender o que está acontecendo nos complexos processos pelos quais as decisões são tomadas algorítmicamente, a *accountability* torna-se impossível. E é essencial apontar que sistemas de inteligência artificial são concebidos como agentes, dedicados a resolver problemas e realizar tarefas (Russell, 2019).

Os algoritmos não devem ser pensados como caixas-pretas que podem ser abertas e cujo conteúdo secreto aguarda divulgação. Existem camadas de complexidade envolvendo inter-relações imprevisíveis entre algoritmos e entre humanos e máquinas na maneira como a IA funciona, tornando a transparência um conceito um tanto estranho. Não há muito a ser revelado abaixo da superfície porque a superfície está continuamente se desdobrando e se tornando algo diferente conforme os *feedbacks* que estes sistemas socio-técnicos recebem. Mesmo que o desenho de um algoritmo não possa se tornar totalmente transparente e compreensível, uma compreensão mais clara de seus resultados parece mais alcançável.

Isso ajuda a resolver outro problema. Frequentemente, argumenta-se que vários sistemas algorítmicos não podem ser mais transparentes: alguns sistemas são até protegidos por leis de sigilo comercial e, se essas leis forem banidas, podem causar retrocessos nos desenvolvimentos tecnológicos (Pasquale, 2015). Além disso, outros algoritmos precisam permanecer opacos por motivos de segurança nacional. Se as operações de alguns sistemas de inteligência artificial fossem totalmente divulgadas, criminosos e inimigos poderiam distorcê-los para servir a seus propósitos, manipulando esses sistemas de maneira a impedir a realização das tarefas para as quais foram projetados. No entanto, as decisões também não poderiam ser totalmente transparentes em uma era de ouro imaginária da ação humana: “Na verdade, os tomadores de decisão humanos são, em última análise, opacos; não há maneiras de rastrear e discernir seus processos psicológicos

‘internos’ usados para chegar a uma decisão” (Bryson, 2021).

Se nos concentrarmos nos resultados das decisões – analisando os algoritmos em vez de seu design – o problema pode ser parcialmente atenuado. A dificuldade de gerar algumas formas de transparência algorítmica não deve implicar na impossibilidade de tornar atores e organizações *accountables* aos cidadãos. König e Wenzelburger (2021) defendem esse ponto quando argumentam que “a legitimidade da tomada de decisão não está mais enraizada em valores e procedimentos internalizados, mas deve estar ligada à aceitabilidade dos resultados produzidos”.

Dissemos, no entanto, que o foco nos resultados é apenas uma solução parcial. Há outra camada do problema que envolve a decisão do que deve ser protegido e o que deve ser divulgado. O fato de que o desenho de alguns sistemas deve ser protegido não significa que todos os projetos devem ser sempre secretos. A decisão sobre proteger o desenho de um sistema deve ser tomada de forma transparente e clara para a sociedade. A definição do que deve ser tornado público está no cerne das democracias, a maioria das quais enfrenta sistematicamente o *trade-off* entre sigilo e transparência. Sobre esse dilema, a definição do que deve permanecer secreto é política e deve, portanto, ser assumida por públicos democráticos (Chambers, 2004). Alguns segredos são perfeitamente compatíveis com as democracias, e o que os torna legítimos é a construção de processos que justifiquem publicamente sua necessidade.

À guisa de conclusão - Mais ética produz melhor regulação?

Em instituições regulatórias emergentes sobre inteligência artificial, coloca-se em uníssono a ideia de que necessitamos de mais ética para controlar a ação derivada de algoritmos de *machine learning*. A proposta de regulação da inteligência artificial no Brasil, por exemplo, produz um desenho de instituições regulatórias baseados na ideia de que a ética deve ser um padrão de comportamento incentivado para desenvolvedores e que a transparência é uma condição necessária nesse quadro ético mais amplo. O desenho da regulação concentra-se na escolha de um marco institucional voltado para o reconhecimento das falhas e das categorias de riscos que emergem com inteligências artificiais. Essas amplas categorias incluem evidências inconclusivas que levam a ações injustificadas, evidências inescrutáveis que levam à opacidade, evidências equivocadas que levam a vieses indesejados, resultados injustos que levam à discriminação, efeitos transformadores que levam à autonomia e desafios de privacidade informacional e rastreabilidade que levam à responsabilidade moral (Tsamados *et al*, 2022).

Frente aos riscos emergentes de algoritmos de *machine learning*, em particular, *policymakers* pontuam a ética como uma solução que molde o desenvolvimento de tecnologias, incentivando comportamentos e práticas de desenvolvedores como resposta aos riscos emergentes. Embora a ética seja uma condição necessária para alcançar um marco regulatório adequado para a inteligência artificial, ela não é uma condição suficiente para mitigar os riscos envolvidos no seu desenvolvimento. A defesa de uma concepção ética termina por reproduzir posições da indústria de tecnologia relacionadas com uma perspectiva de autorregulação da inteligência artificial (Papyshev; Yarime, 2022). Necessitamos lembrar que o projeto de lei 21/2020 foi aprovado na Câmara dos Deputados com a perspectiva de uma inteligência artificial ética, baseada em processos de autorregulação criados pela própria indústria (Câmara dos Deputados, 2020). Se considerarmos a forma como algoritmos de *machine learning* são construídos para resolver problemas em governos, percebe-se claramente que uma concepção de autorregulação não é suficiente para mitigar os riscos.

A governança de inteligência artificial, especialmente quando aplicada em governos, deve estar baseada em uma arquitetura institucional que valorize a transparência e a *accountability* como princípios essenciais (Filgueiras, 2022). A transparência algorítmica é o elemento pelo qual os cidadãos podem saber como os sistemas de decisão fazem escolhas, com quais propósitos, que impactam suas vidas (Diakopoulos 2016; Hildebrandt 2012). A transparência algorítmica é desafiadora e implica em diferentes condições cognitivas entre designers de sistemas e usuários afetados por esses sistemas. A transparência algorítmica tem limites nas possibilidades de sigilo – tanto em governos quanto em empresas – e nas capacidades cognitivas do cidadão para criar entendimento (Crain, 2018; Ananny; Crawford 2018). Essas diferenças cognitivas implicam desafios

para a transparência dos sistemas de IA, exigindo que não seja o pleno “abrir a caixa preta”, mas processos de justificação das decisões no processo de adoção da IA que criem uma percepção de legitimidade por parte dos usuários e afetados pelas ações dos sistemas sociotécnicos (Fine Licht e Fine Licht, 2020). A *accountability* aplicada a algoritmos está sujeita a controvérsias e escrutínio sobre os valores e princípios envolvidos; requer permanentemente a justificativa de ser usado para o bem público (Binns 2018).

Assim, governos podem superar barreiras de transparência de algoritmos de *machine learning* concentrando-se não necessariamente na sua dinâmica de desenho, que sempre será baseada em tentativa e erro, mas nos resultados de sistemas algorítmicos. Governos necessitam criar arquiteturas institucionais que possibilitem reduzir os dilemas de delegação do processo decisório para agentes inteligentes baseados em algoritmos de *machine learning*. Estas arquiteturas institucionais que estabeleçam a transparência da tomada de decisão e realização de tarefas, operadas por sistemas baseados em *machine learning*, devem considerar os seguintes aspectos:

- Tornar as bases de dados de treinamento abertas, quando apropriado, para escrutínio público, de forma que as soluções possam ser testadas e reproduzidas pela sociedade;
- Relatar os testes de acurácia dos modelos empregados em algoritmos de *machine learning*, de maneira a informar a sociedade sobre as escalas alcançadas e precisão da informação produzida;
- Relatar como as bases de dados de *feedback* são construídas e processadas e empregadas em plataformas ou sistemas;
- Estabelecer mecanismos de supervisão humana dos resultados e do trabalho de sistemas sociotécnicos baseados em algoritmos de *machine learning*;
- Transparecer resultados, quando apropriado, de forma a dar *feedback* à sociedade do processo decisório baseado em algoritmos de *machine learning*;
- Relatar para a sociedade aspectos relacionados à qualidade dos dados utilizados em sistemas baseados em *machine learning* e atividades relacionadas para a detecção e eliminação de vieses;
- Relatar os riscos no processo decisório de políticas quando realizados por sistemas;
- Relatar a consistência do sistema com o desenho e teoria da política ou programa público;
- Criação de uma nova estrutura de direitos digitais que permitam a cidadãos requerer, contestar ou pedir revisão de decisões tomadas por sistemas de *machine learning*.

Criar arquiteturas institucionais é essencial para superar o desafio da transparência quando adicionamos algoritmos de *machine learning* como ferramentas para solução de problemas. Os tópicos levantados acima dependem de arquiteturas que sejam moldadas em novas práticas de gestão e processos de governança da inteligência artificial, especialmente enfocadas em governos e na administração pública. Como se trata de novas demandas para a administração, processos de tentativa e erro permanecerão no horizonte, tendo em vista decisões incrementais que vão aprimorando o processo. A adoção de algoritmos de *machine learning* representa importantes inovações governamentais. Superar os desafios de transparência que essa tecnologia cria para governos significa estabelecer um critério de interesse público a instituições algorítmicas emergentes, com especial foco na sua democratização e publicidade no contexto de sociedades cada vez mais epistemicamente transformadas por sistemas e suas aplicações.

Referências

ABERS, Rebecca. **Ativismo institucional: Criatividade e luta na burocracia brasileira**. Brasília: Editora UnB, 2021.

AMOORE, Louise. Machine learning political orders. **Review of International Studies**, v.49, n.1, p. 20-36, 2022. DOI: <https://doi.org/10.1017/S0260210522000031>.

AMOORE, Louise. The deep border. **Political Geography**, v.96, p.1-9, 2021. DOI: <https://doi.org/10.1016/j.polgeo.2021.102547>.

ANANNY, Mark; CRAWFORD, Kate. Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. **New Media & Society**, v.20, n.3, p. 973–989, 2018. DOI: <https://doi.org/10.1177/1461444816676645>.

ANSELL, Chris. Ecological explanation. In: BERK, G.; GALVAN, D.C.; HATTAM, V. (org.). **Political creativity**. Reconfiguring institutional order and change. Philadelphia: University of Pennsylvania Press, 2013. p. 55–77.

BENJAMIN, Ruha. **Race after technology**. New York: Polity Press, 2019.

BINNS, Reuben. Algorithmic accountability and public reason. **Philosophy & Technology**, v.31, n.4, p. 543–556, 2018. DOI: <https://doi.org/10.1007/s13347-017-0263-5>.

BRYSON, Joanna. **AI & global governance: No one should trust AI**, 2018. Disponível em: <https://cpr.unu.edu/publications/articles/ai-global-governance-no-one-should-trust-ai.html>. Acesso em: 07 out 2023.

CÂMARA DOS DEPUTADOS. **Projeto de Lei 21/2020**. Brasília: Câmara dos Deputados. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 07 out 2023.

CHAMBERS, Simone. Behind closed doors: Publicity, secrecy, and the quality of deliberation. **Journal of Political Philosophy**, v.12, p.389-410. DOI: <https://doi.org/10.1111/j.1467-9760.2004.00206.x>.

COECKELBERGH, Mark. Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. **AI & Ethics**, early view, 2022. DOI: <https://doi.org/10.1007/s43681-022-00239-4>.

COLLINGTON, Rosie. Disrupting the Welfare State? Digitalisation and the Retrenchment of Public Sector Capacity. **New Political Economy**, v.27, p.2, p.312-328, 2022. DOI: <https://doi.org/10.1080/13563467.2021.1952559>.

CRAIN, Matthew. The limits of transparency: data brokers and commodification. **New Media & Society**, v.20, n.1, p. 88–104, 2018. DOI: <https://doi.org/10.1177/1461444816657096>.

DAI, Xi. Enforcing law and norms for good citizens: One view of China’s social credit system project. **Development**, v.63, p. 38-43, 2020. DOI: <https://doi.org/10.1057/s41301-020-00244-2>.

DIAKOPOULOS, Nick. Accountability in algorithmic decision making. **Communications of the ACM**, v.59, p. 2, p.56–62, 2016. DOI: <https://doi.org/10.1145/2844110>.

EISENHARDT, Kathleen M. Agency Theory: An Assessment and Review. **The Academy of Management Review**, v.14, n.1, p.57–74, 1989. DOI: <https://doi.org/10.5465/amr.1989.4279003>

EUBANKS, Virginia. **Automating inequality**. How high-tech tools profile, police, and punish the poor. New York: St. Martin’s Press, 2018.

FILGUEIRAS, Fernando. New Pythias of public administration: Ambiguity and choice in AI systems and challenges for governance. **AI & Society**, v.37, n.4, p.1473-1486, 2022. DOI: <https://doi.org/10.1007/s00146-021-01201-4>.

FILGUEIRAS, Fernando. The silent reform: Digital governance as a strategy for state reform in Brazil.

In: LISBOA, Erika; GOMES, Ricardo C.; MARTINS, Humberto F. (eds.). **The Brazilian way of doing public administration**. Bingley: Emerald, 2023. p. 83-95. DOI: <https://doi.org/10.1108/978-1-80262-655-120231008>

FILGUEIRAS, Fernando. Transparency and accountability: Principles and rules for the construction of publicity. **Journal of Public Affairs**, v.16, n.2, p.192-202, 2016. DOI: <https://doi.org/10.1002/pa.1575>.

FINE LICHT, Karl de; FINE LICHT, Jenny de. "Artificial intelligence, transparency, and public decision-making". **AI & Society**, 35 (4), 917–926, 2020. DOI: <https://doi.org/10.1007/s00146-020-00960-w>

FRISCHMANN, Brett; SELINGER, Evan. **Re-engineering humanity**. Cambridge: Cambridge University Press, 2018.

GAILMARD, Sean. Multiple principals and oversight of bureaucratic policy-making". **Journal of Theoretical Politics**, v.21, n. 2, p. 161–86, 2009. DOI: <https://doi.org/10.1177/0951629808100762>.

GORWA, Robert; BINNS, Reuben, KATZENBACH, Christian. Algorithmic content moderation: Technical and political challenges in automation of platform governance. **Big Data & Society**, v.7, n.1, p.1-15, 2020. DOI: <https://doi.org/10.1177/2053951719897945>.

INNERARITY, Daniel. Making the black box society transparent. **AI & Society**, v.36, 975–981, 2021. DOI: <https://doi.org/10.1007/s00146-020-01130-8>.

ISSAR, Shiv; ANEESH, Aneesh. What is algorithmic governance? **Sociology Compass**, v.16, n. 1, e12955, 2022. DOI: <https://doi.org/10.1111/soc4.12955>.

JONES, Bryan D.; BAUMGARTNER, Frank R. **The politics of attention: How government prioritizes problems**. Chicago: The University of Chicago Press, 2005.

KELLEHER, J. D. **Deep learning**. Cambridge: The MIT Press, 2019.

KNUTH, D. E. **The art of computer programming**. Volume 1: Fundamental Algorithms. Berkeley: Addison-Wesley, 1968.

KÖNIG, P. D.; WENZELBURGER, G. The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it. **Technology in Society**, v. 67, p.1-10, 2021. DOI: <https://doi.org/10.1016/j.techsoc.2021.101688>.

KROLL, Joshua A.; HUEY, Joanna; BAROCAS, Solon; FELTEN, Edward W.; REIDENBERG, Joel R.; ROBINSON, David G.; YU, Harlan. (2017). Accountable algorithms. **University of Pennsylvania Law Review**, v.165, p.633–705. Disponível em: https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3.

LINDBLOM, Charles. The science of muddling through. **Public Administration Review**, v.19, n. 2, p.79-88. DOI: <https://doi.org/10.2307/973677>.

MARKRAM, Henry. The Blue Brain project. **Nature Reviews Neuroscience**, v. 7, n. 2, p. 153–160, 2006.

MENDONÇA, Ricardo F.; FILGUEIRAS, Fernando; ALMEIDA, Virgílio A. **Algorithmic Institutionalism**. The Changing Rules of Social + Political Life. Oxford: Oxford University Press, 2023.

MINSKY, Marvin. **The Society of Mind**. New York: Simon and Schuster, 1985.

OBERMEYER, Ziad; POWERS, Brian; VOGELI, Christine; MULLAINATHAN, Sendhil. Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, v.366 , n.6464, 447–453, 2019. DOI: <https://doi.org/10.1126/science.aax2342>.

PAPYSHEV, Gleb; YARIME, Masaru. The limitation of ethics-based approaches to regulating artificial intelligence: regulatory gifting in the context of Russia. **AI & Society**, early view, p. 1-16. DOI: <https://doi.org/10.1007/s00146-022-01611-y>.

PASQUALE, Frank. **The black box society**. The secret algorithms that control money and information. Cambridge: Harvard University Press, 2015.

RUSSELL, Stuart J. **Human compatible**: Artificial intelligence and the problem of control. New York: Viking Books, 2019.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial intelligence**: A modern approach. Englewood Cliffs, NJ: Prentice-Hall, 2010.

SAMUEL, A.L. Artificial intelligence: A frontier of automation. **The Annals of the American Academy of Political and Social Science**, v.340, n.1, p.10-20, 1962. DOI: <https://doi.org/10.1177/000271626234000103>.

SAMUEL, Arthur. Some studies in machine learning using the game of checkers. **IBM Journal of Research and Development**, v. 44, n. 1-2, p. 210–229, 1959. DOI: <https://doi.org/10.1147/rd.33.0210>.

SIMON, Herbert A. **Models of man**: Social and rational. New York: John Wiley, 1957.

SIMON, Herbert A. **The sciences of the artificial**. Cambridge, MA: MIT Press, 1969.

SIMON, Herbert A. Artificial intelligence: An empirical science. **Artificial Intelligence**, v.77, n.1, p.95-127. DOI: [https://doi.org/10.1016/0004-3702\(95\)00039-H](https://doi.org/10.1016/0004-3702(95)00039-H).

TSAMADOS, Andreas; AGGARWAL, Nikita; COWLS, Josh; MORLEY, Jessica; ROBERTS, Huw; TADDEO, Mariarosaria; FLORIDI, Luciano. The ethics of algorithms: key problems and solutions. **AI & Society**, v.37, p.1, p. 215–230, 2022. DOI: <https://doi.org/10.1007/s00146-021-01154-8>.

TURING, Alan. Computing machinery and intelligence. **Mind—A Quarterly Review of Psychology and Philosophy**, v.59, n.236, p.433–460, 1950.

WANG, Pei. On defining artificial intelligence. **Journal of Artificial General Intelligence**, v.10 , n.2, p.1-37, 2019. DOI: <https://doi.org/10.2478/jagi-2019-0002>.

Recebido em 12 de abril de 2023.
Aceito em 16 de maio de 2023.