


**CIÊNCIA DE DADOS COMO INFRAESTRUTURA EPISTÊMICA DO ESTADO:  
GOVERNANÇA DIGITAL E VALORES REPUBLICANOS EM POLÍTICAS  
PÚBLICAS**


**DATA SCIENCE AS THE EPISTEMIC INFRASTRUCTURE OF THE STATE:  
DIGITAL GOVERNANCE AND REPUBLICAN VALUES IN PUBLIC POLICIES**

Recebido em: 12/03/2026

Aceito em: 27/03/2026

Publicado em: 16/04/2026

Marcelo Salvador Celestino<sup>1</sup>   
Universidade Estadual Paulista

Vânia Cristina Pires Nogueira Valente<sup>2</sup>   
Universidade Estadual Paulista

**Resumo:** O ensaio analisa a integração entre Ciência de Dados, infraestrutura analítica e Governança Digital na Administração Pública, enfatizando a precedência dos valores republicanos e a legitimidade democrática nas decisões orientadas por evidências. Propõe-se compreender a Ciência de Dados como infraestrutura epistêmica do Estado, responsável por estruturar a produção de evidências e ampliar a capacidade analítica da ação estatal. Nesse enquadramento, valores públicos, governança institucional e infraestrutura técnica constituem dimensões interdependentes que condicionam a legitimidade das decisões orientadas por dados. O estudo examina riscos associados à racionalidade algorítmica, como a autonomização da técnica, a opacidade decisória e a reprodução de desigualdades, e analisa mecanismos de mediação institucional baseados em Governança Digital, supervisão humana e contestabilidade. Casos documentados e um exercício analítico ilustram como a subordinação da técnica a valores republicanos pode orientar decisões públicas legítimas. Ao articular fundamentos conceituais, arquitetura analítica e implicações institucionais, o ensaio contribui para o debate sobre Políticas Públicas digitais e para a compreensão das capacidades estatais em contextos de Governança Algorítmica.

**Palavras-chave:** Administração Pública; Ciência de Dados; Inteligência Artificial; Governança Digital; Políticas Públicas Baseadas em Evidências.

**Abstract:** This essay analyzes the integration of Data Science, analytical infrastructure, and Digital Governance in Public Administration, emphasizing the precedence of republican values and democratic legitimacy in evidence-informed decision-making. It proposes understanding Data Science as the State's epistemic infrastructure, responsible for structuring the production of evidence and expanding the analytical capacity of governmental action. Within this framework, public values, institutional governance, and technical infrastructure constitute interdependent dimensions that condition the legitimacy of data-oriented decisions. The study examines risks associated with algorithmic rationality, including the autonomization of technique, decision opacity, and the reproduction of inequalities, and discusses institutional mediation mechanisms based on Digital Governance, human oversight, and contestability. Documented cases and an analytical exercise illustrate how subordinating technical systems to republican values can guide legitimate public decisions. By articulating conceptual foundations, analytical architecture, and institutional implications, the essay contributes to debates on digital Public Policies and to the understanding of state capacities in contexts of Algorithmic Governance.

**Keywords:** Artificial Intelligence; Data Science; Digital Governance; Evidence-Informed Public Policies; Public Administration.

---

<sup>1</sup> Doutor e Mestre pelo Programa de Pós-Graduação em Mídia e Tecnologia (PPGMiT), Faculdade de Arquitetura, Artes, Comunicação e Design (FAAC), Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP). Bacharel em Administração, Bauru, SP, Brasil. E-mail: marcelo.salvador@unesp.br

<sup>2</sup> Livre-Docente em Representação Gráfica pela Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) e Doutora em Engenharia Civil pela Universidade de São Paulo (POLI-USP). Professora Associada e Coordenadora do Programa de Pós-Graduação em Mídia e Tecnologia (PPGMiT-DO), Faculdade de Arquitetura, Artes, Comunicação e Design (FAAC) (UNESP), Bauru, SP, Brasil. E-mail: vania.valente@unesp.br

## INTRODUÇÃO

A expansão da Governança na Era Digital (GED) tem transformado profundamente as formas de produção e uso de informações no setor público. Dunleavy e Margetts (2023) descrevem esse processo como uma terceira onda da governança digital, caracterizada pela substituição da simplificação administrativa de dados — típica das burocracias tradicionais — por arquiteturas digitais capazes de processar grandes volumes de informações em tempo real. Nesse novo ambiente institucional, registros administrativos integrados, bases de dados interoperáveis e sistemas de Inteligência Artificial (IA) ampliam a capacidade analítica do Estado, permitindo diagnósticos mais precisos e intervenções públicas mais informadas.

A consolidação dessas infraestruturas digitais tem sido discutida na literatura de Administração Pública como parte da transformação informacional do Estado. Bekkers, Duivenboden e Thaens (2006) destacam que o uso sistemático de dados e registros administrativos integrados altera a forma como governos produzem conhecimento sobre problemas públicos, ampliando as capacidades de monitoramento, coordenação e formulação de políticas. Nessa perspectiva, a infraestrutura informacional do Estado torna-se elemento central para a produção de diagnósticos e para a sustentação técnica das decisões governamentais (Jannuzzi, 2009, 2013).

Essa transformação também dialoga com debates mais amplos sobre capacidade estatal nas ciências sociais. De modo geral, capacidade estatal refere-se à habilidade institucional de formular e implementar Políticas Públicas de maneira eficaz, coordenando recursos administrativos, conhecimento técnico e instrumentos regulatórios (Cingolani, 2013; Fukuyama, 2013). A incorporação de infraestruturas digitais, bases de dados integradas e sistemas de análise avançada configura uma dimensão emergente dessa capacidade, na qual a produção sistemática de evidências passa a desempenhar papel central na ação governamental.

Nesse contexto, a Política Baseada em Evidências (PBE), inspirada na Medicina Baseada em Evidências (Sackett *et al.*, 1996), ampliou o uso de indicadores, métricas e métodos analíticos na formulação de políticas. Embora esses instrumentos contribuam para qualificar decisões públicas, sua adoção também levanta questões relevantes sobre legitimidade democrática e autoridade técnica. Como observa Jannuzzi (2009, 2018), indicadores e métricas podem deslocar o debate político para o plano técnico quando operam sem parâmetros normativos explícitos, conferindo aparência de neutralidade a decisões que envolvem escolhas metodológicas e preferências analíticas.

Essa tensão entre capacidade analítica e legitimidade normativa torna-se particularmente relevante na Administração Pública contemporânea, marcada pela crescente

centralidade dos dados e dos sistemas algorítmicos. A expansão dessas infraestruturas não apenas amplia a capacidade do Estado de produzir evidências, mas também reconfigura os critérios pelos quais problemas públicos são identificados, interpretados e priorizados.

Partindo dessa problemática, este ensaio propõe compreender a Ciência de Dados como *infraestrutura epistêmica* do Estado. Nessa perspectiva, as estruturas analíticas que organizam dados, indicadores e algoritmos não atuam apenas como instrumentos técnicos de apoio à decisão, mas como dispositivos institucionais que moldam os modos de conhecer, interpretar e governar a realidade social. A noção de infraestrutura epistêmica estatal é, assim, apresentada como categoria analítica para compreender a autonomia informacional do Estado em contextos de crescente dependência de sistemas digitais.

Como referencial analítico, mobiliza-se o modelo de Políticas Públicas Baseadas em Valores e Informadas por Evidências (PPVE), proposto por Jannuzzi (2024a, 2024b, 2025). Esse modelo permite examinar a articulação entre produção de evidências, Governança Digital e valores republicanos, enfatizando que o uso de evidências na formulação de políticas deve permanecer subordinado a parâmetros normativos explícitos e a mecanismos institucionais de supervisão e responsabilização.

Diferentemente de abordagens predominantes sobre governo orientado por dados, frequentemente associadas a foco em eficiência administrativa ou mecanismos regulatórios (Dunleavy; Margetts, 2023), este ensaio propõe uma leitura institucional da infraestrutura analítica do Estado. Em vez de tratar dados e algoritmos apenas como instrumentos de decisão, o artigo os interpreta como componentes de uma infraestrutura epistêmica, cuja legitimidade depende da mediação normativa da Governança Digital e da precedência de valores republicanos.

## ESTUDOS CORRELATOS

A literatura recente tem investigado diferentes dimensões da Gestão Digital e da legitimidade democrática associada ao uso de sistemas algorítmicos na ação governamental. Monteiro e Marrafon (2024) propõem parâmetros iniciais para a aplicação da IA em políticas orientadas por dados, apresentando um modelo tridimensional no qual a legitimidade das decisões depende da representatividade dos atores envolvidos (*input*), da qualidade dos processos decisórios e do devido processo democrático (*throughput*) e da justiça social dos resultados produzidos (*output*).

Na mesma direção, Mark e Morison (2025) analisam estratégias institucionais destinadas a equilibrar inovação tecnológica e *accountability* (responsabilização e prestação de

contas) em decisões mediadas por IA. Os autores alertam para uma tendência internacional de flexibilização regulatória — descrita como um processo de “abrandamento regulatório” — no qual a busca por competitividade geopolítica e tecnológica pode reduzir os mecanismos de supervisão humana e controle institucional sobre sistemas algorítmicos.

No contexto brasileiro, a tese de Santos (2024) examina a transição para o Estado Digital, descrevendo a formação de um “Microsistema de Transição Digital”, no qual a Governança de Dados passa a desempenhar papel central na qualificação da tomada de decisão pública, substituindo práticas baseadas predominantemente em julgamento intuitivo por processos orientados por evidências estruturadas, em linha com a perspectiva de Jannuzzi (2018) sobre avaliação de Políticas Públicas, que ressalta a mediação normativa e a redução da discricionariedade individual em decisões complexas.

Apesar desses avanços, permanecem relativamente escassos os estudos que articulam de forma explícita valores republicanos, Governança Digital e a infraestrutura epistêmica associada à Ciência de Dados como dimensões integradas de um mesmo arranjo institucional. É nesse ponto que o presente ensaio busca contribuir, ao propor uma interpretação institucional da infraestrutura analítica do Estado que enfatiza a mediação normativa dos valores públicos na produção e no uso de evidências.

## **O MODELO DE POLÍTICAS PÚBLICAS BASEADAS EM VALORES E INFORMADAS POR EVIDÊNCIAS**

No presente debate, destaca-se o modelo de PPVE, de Jannuzzi (2024a, 2024b, 2025), que propõe integrar diferentes fontes de conhecimento no processo decisório público. O modelo enfatiza que o uso de evidências na formulação de políticas deve permanecer subordinado a valores públicos e ao arcabouço institucional do Estado, incorporando não apenas evidências científicas e registros administrativos, mas também conhecimento técnico, experiências de implementação e contextos institucionais específicos.

Nesse enquadramento, a produção e o uso de evidências não se organizam em um processo linear, mas em uma dinâmica de retroalimentação contínua entre formulação, implementação e avaliação, descrita pelo autor como uma espiral de implementação, na qual novos dados e aprendizados institucionais contribuem para o aprimoramento progressivo das Políticas Públicas. Em formulações mais recentes, o autor também reconhece o papel crescente das infraestruturas digitais e das informações algorítmicas na produção e circulação de evidências no campo de públicas, destacando a necessidade de mediações institucionais

capazes de alinhar essas capacidades analíticas aos princípios normativos da ação pública (2024a, 2024b, 2025).

## **GOVERNANÇA DE DADOS E LEGITIMAÇÃO INSTITUCIONAL**

A Governança de Dados orienta o uso legítimo da base analítica na Administração Pública, enfatizando transparência, explicabilidade e supervisão humana como condições essenciais de autoridade normativa. Ainda assim, a base normativa que sustenta decisões orientadas por evidências permanece pouco explorada, evidenciando a necessidade de debater seus aspectos éticos e sociais antes que esses sistemas se tornem ubíquos e invisíveis (Coeckelbergh, 2020). Nesse contexto, a governança estabelece os critérios institucionais para o uso de dados, definindo quem os determina e quais valores orientam essa estrutura.

A criação de ecossistemas informacionais integrados amplia a capacidade diagnóstica e prospectiva do Estado, ao mesmo tempo em que reorganiza relações de poder e redefine as fronteiras da Administração Pública. A Ciência de Dados passa a atuar como motor da racionalidade administrativa. Nesse cenário, a lógica da eficiência (racionalidade funcional) deve ser equilibrada pela lógica da adequação, fundamentada em valores políticos e legais legitimados (racionalidade substancial) (Bekkers; Duivenboden; Thaens, 2006). No Brasil, a interação entre dados abertos, PBE e proteção de dados ilustra essa tensão: enquanto a abertura fortalece a transparência e o controle social, a centralidade da medição e da quantificação pode deslocar o juízo político para métricas apresentadas como neutras.

Nesse cenário, a Governança de Dados públicos estrutura-se pela coexistência de duas agendas complementares: a transparência informacional, promovida pela Lei de Acesso à Informação (Lei nº 12.527/2011) e pela Política de Dados Abertos do Poder Executivo; e a proteção de dados pessoais, estabelecida pela Lei Geral de Proteção de Dados Pessoais (LGPD) (Lei nº 13.709/2018). Enquanto a abertura de dados fortalece a transparência e o controle social, a regulação da privacidade impõe limites ao tratamento informacional, exigindo mecanismos institucionais capazes de equilibrar segurança e responsabilização (Brasil, 2011; Brasil, 2018).

A LGPD estabelece princípios como finalidade, adequação, necessidade, livre acesso, qualidade, transparência, segurança, prevenção, não discriminação e responsabilização (Brasil, 2018). Esses princípios orientam a coleta, o armazenamento e o uso de dados pessoais, assegurando que o tratamento seja legítimo, proporcional e transparente, além de garantir aos titulares controle efetivo sobre suas informações. Diretrizes internacionais, como as da Organisation for Economic Co-operation and Development (OECD) (2001), reforçam a

necessidade de limites à coleta, padrões de segurança informacional e mecanismos claros de responsabilização do controlador.

Apesar desses marcos normativos, decisões automatizadas — especialmente aquelas baseadas em aprendizado profundo (*deep learning*) — permanecem frequentemente opacas, evidenciando uma assimetria epistêmica entre sistemas técnicos e julgamento humano (Coeckelbergh, 2020). A governança jurídica cria instrumentos de contestação e auditoria, mas não elimina a dependência de infraestruturas técnicas complexas. Por isso, a integração consistente de dados, a padronização de metadados e a *accountability* interinstitucional tornam-se fundamentais para garantir coerência informacional e legitimidade decisória.

A intensificação do uso de dados exige rigor nos processos de integração informacional. Falhas em etapas de tratamento e carga de dados podem cristalizar vieses e comprometer a integridade analítica dos sistemas, podendo resultar em impactos coletivos incalculáveis. Por isso, a padronização de metadados e a rastreabilidade dos fluxos informacionais tornam-se centrais para assegurar consistência decisória e responsabilização institucional. A legitimidade de decisões automatizadas depende não apenas de transparência formal, mas sobretudo da capacidade de explicação pública dos critérios utilizados. Como observa Coeckelbergh (2020), a explicabilidade é fator determinante para a confiança social em sistemas de IA, enquanto a simples abertura de código é insuficiente sem mecanismos efetivos de responsabilização.

Nesse quadro analítico, ampliar a base cognitiva da decisão pública também se mostra relevante. O “igualitarismo epistêmico” sustenta que decisões coletivas podem ser fortalecidas ao mobilizar recursos de conhecimento distribuídos na sociedade, e não apenas entre especialistas ou elites técnicas (Allen, 2023). De modo semelhante, a “governança inteligente” propõe conectar instituições públicas a redes ampliadas de especialistas e cidadãos capazes de contribuir com diferentes formas de *expertise* (perícia) na resolução de problemas públicos (Noveck, 2015).

Contudo, a ampliação dessas bases cognitivas não dispensa mediação normativa, já que rigor metodológico, por si só, não assegura legitimidade. Evidências tornam-se justificáveis apenas quando interpretadas à luz de princípios constitucionais explicitamente definidos. Portanto, a Governança de Dados pode ser compreendida como mediação institucional entre capacidade analítica e validade institucional, enquanto a evidência permanece central na decisão pública, mas subordinada a finalidades normativas republicanas.

## INFRAESTRUTURA ANALÍTICA E PRODUÇÃO DE EVIDÊNCIAS NA ADMINISTRAÇÃO PÚBLICA

A infraestrutura analítica organiza a produção de evidências na Administração Pública contemporânea. Sua qualidade depende não apenas da disponibilidade de dados, mas também da arquitetura institucional que orienta sua produção, integração e validação. As Bases de Dados Governamentais (BDG) estruturam informações administrativas, fiscais e sociais em Sistemas de Monitoramento e Avaliação (SM&A) destinados a subsidiar a tomada de decisão e aprimorar a ação pública. Nesses arranjos, a evidência é produzida em sistemas permanentes que integram informações de múltiplas fontes, conferindo-lhes significado analítico e assegurando sua disponibilidade oportuna para subsidiar decisões públicas (tempestividade) (Jannuzzi, 2013).

A expansão dessa infraestrutura analítica reflete o amadurecimento do poder infraestrutural do Estado, definido por Mann (1984) como a capacidade institucional de penetrar na sociedade civil e implementar decisões logísticas em todo o território. Diferentemente do poder despótico — baseado em ações arbitrárias da elite sem negociação — o poder infraestrutural se manifesta na capacidade do Estado de atuar como um ponto de condensação e coordenação das relações sociais, articulando instituições, informações e capacidades administrativas.

Entretanto, Ellul (1980) chama atenção para o fato de que o desenvolvimento do “sistema técnico” tende a deslocar progressivamente a centralidade da decisão política. À medida que a técnica se organiza em um sistema integrado e autoexpansivo, as escolhas passam a ser apresentadas como imperativos técnicos de eficiência e racionalidade. Nesse contexto, os responsáveis políticos deixam de definir finalidades coletivas e passam a atuar sobretudo como mediadores ou legitimadores de decisões já estruturadas pela lógica técnica, convertendo-se, muitas vezes, em porta-vozes das soluções produzidas pelo próprio sistema tecnológico.

Historicamente, essa capacidade foi impulsionada por técnicas de armazenamento e recuperação de informação, como a alfabetização e a estatística, que permitiram codificar responsabilidades administrativas e organizar dados de forma sistemática. Na contemporaneidade, as BDG representam uma evolução dessas técnicas, permitindo que o Estado armazene e recupere volumes massivos de informação sobre a população. Assim, a infraestrutura informacional não apenas subsidia decisões públicas, mas também territorializa a vida social ao estruturar fluxos de informação que sustentam a ação estatal (Mann, 1984).

A organização técnica dessas bases é fundamental para a consistência da evidência pública. Estruturas relacionais e princípios de integridade lógica garantem que a informação seja representada de forma padronizada, com valores explícitos em colunas e linhas de tabelas interpretadas como proposições verificáveis (Date; Kannan; Swamynathan, 2006). O modelo relacional, fundamentado em lógica e matemática, fornece o rigor necessário para que sistemas de dados operem de maneira previsível e consistente. Nesse contexto, a Linguagem de Consulta Estruturada (*Structured Query Language*, SQL) consolida-se como padrão internacional para manipulação e recuperação de dados, permitindo examinar tanto grandes conjuntos de informação quanto registros individuais (Beaulieu, 2010).

A conversão de registros administrativos em evidência analítica exige ferramentas especializadas de Ciência de Dados. Bibliotecas da linguagem *Python*, como *NumPy*, oferecem funções aritméticas e estruturas eficientes para processamento científico, enquanto a biblioteca *Pandas* possibilita manipulação flexível de estruturas tabulares (*DataFrames*) para análise exploratória e preparação de dados (Feltrin, 2021). Complementarmente, o conjunto de pacotes *Tidyverse*, na linguagem *R*, enfatiza a organização sistemática dos dados (*data tidying*), estruturando-os de modo que cada coluna represente uma variável e cada linha uma observação, facilitando etapas posteriores de modelagem e visualização (Wickham; Çetinkaya-Rundel; Grolemond, 2023).

Esse processo ocorre por meio de rotinas de Extração, Transformação e Carregamento (*Extract, Transform, Load*, ETL), que estruturam fluxos de coleta, tratamento e armazenamento de dados, permitindo superar o paradoxo da “escassez na abundância” por meio de procedimentos rigorosos de validação e integração informacional (Jannuzzi, 2013). As etapas do ETL podem ser sintetizadas da seguinte forma: 1) Extração – coleta de dados provenientes de diferentes sistemas e bases administrativas, integrando-os segundo unidades de referência comuns, como identificadores individuais ou territoriais; 2) Transformação – limpeza, padronização e harmonização dos dados (*wrangling*), etapa crítica para tratar valores ausentes, inconsistências e erros operacionais que podem comprometer a integridade analítica das bases (Feltrin, 2021; Wickham; Çetinkaya-Rundel; Grolemond, 2023); 3) Carregamento – armazenamento estruturado em repositórios analíticos, como *data warehouses*, para dados estruturados, ou *data lakes*, para dados brutos, tornando os dados persistentes e acessíveis para análise, monitoramento e reprogramação de Políticas Públicas (Date; Kannan; Swamynathan, 2006; Jannuzzi, 2013).

A Imagem 1 ilustra, de forma esquemática (e simbólica), o processo pelo qual dados administrativos inicialmente dispersos são convertidos em evidência estruturada, capaz de apoiar o planejamento, o monitoramento e a avaliação de Políticas Públicas.

Imagem 1 – Representação das fases de extração, transformação e carga dos dados.



**Fonte:** Elaborado pelos autores (2026), com auxílio da ferramenta de Inteligência Artificial generativa Nano Banana 2 (Google). Nota: As pedras representam os dados brutos, que passam por um processo de tratamento até se transformarem em blocos estruturados e úteis para análise.

Após o processo de ETL, os dados passam a alimentar modelagem estatística e aplicações de IA, permitindo capturar o sinal dos fenômenos observados e reduzir o ruído presente nas bases informacionais. Esse processamento amplia a capacidade inferencial do Estado, possibilitando estimativas causais e projeções prospectivas mais robustas (Hastie; Tibshirani; Friedman, 2009; Wickham; Çetinkaya-Rundel; Golemund, 2023). Técnicas como regressões, métodos de regularização e validação cruzada asseguram a generalização dos modelos e reduzem problemas de sobreajuste (*overfitting*), equilibrando viés e variância na estimação (Hastie; Tibshirani; Friedman, 2009).

Quando experimentos aleatorizados não são viáveis por restrições éticas ou operacionais, desenhos quase-experimentais — como pareamento (*matching*) e Diferenças em Diferenças — permitem estimar impactos de Políticas Públicas em condições institucionais reais, exigindo rigor analítico para evitar inferências precipitadas. Essa abordagem busca contrapor-se à “ditadura do método” e ao mito do “padrão-ouro”, reconhecendo que a complexidade da implementação e a heterogeneidade territorial impõem limites à validade dos experimentos clássicos (Jannuzzi, 2018).

Algoritmos de aprendizado de máquina (*machine learning*) ampliam a dimensão prospectiva da ação estatal por meio de técnicas supervisionadas de classificação e regressão (Dunleavy; Margetts, 2023; Santos, 2024). Modelos preditivos permitem identificar riscos, detectar fraudes e antecipar demandas por serviços públicos, inserindo-se em arranjos de

governança orientados por dados que conectam capacidade analítica, eficiência administrativa e responsividade estatal (Dunleavy; Margetts, 2023; Santos, 2024).

A expansão dessas capacidades digitais reforça o poder infraestrutural do Estado, definido como a capacidade institucional de penetrar na sociedade civil e implementar decisões logísticas em todo o território (Mann, 1984). No contexto contemporâneo, o trabalho prévio de inteligência de dados orienta novos rumos na gestão pública e sustenta a legitimidade das decisões estatais por meio de sua infraestrutura informacional (Santos, 2024). O Quadro 1 sintetiza os principais modelos de IA aplicados à gestão pública, destacando seu potencial de geração de valor público e a necessidade de adequação a parâmetros de transparência e justiça social.

Quadro 1 – Modelos de IA aplicados ao Governo.

Modelo de IA	Aplicação	Benefícios	Limitações e Desafios	Ref.
<b>Neural Networks (Redes Neurais Artificiais)</b>	Análise de padrões complexos e não lineares (saúde pública, detecção de anomalias)	Alta capacidade preditiva em grandes volumes de dados	Natureza de “caixa-preta” e baixa interpretabilidade	[1-3]
<b>Random Forests (Florestas Aleatórias)</b>	Identificação de grupos de risco em saúde e arrecadação fiscal	Reduz o risco de sobreajuste	Exige volumes maiores de dados para bom desempenho	[2-4]
<b>Decision Trees (Árvores de Decisão)</b>	Classificação de beneficiários ou riscos tributários	Alta interpretabilidade e transparência nos critérios	Sensível a ruídos e instabilidade em bases pequenas	[2,3]
<b>K-means (Agrupamento k-means)</b>	Segmentação populacional e análise territorial de demandas	Identifica padrões em dados não rotulados (aprendizado não supervisionado)	Sensível à definição prévia do número de agrupamentos ( <i>k</i> )	[2,3]
<b>Support Vector Machine (Máquinas de Vetores de Suporte)</b>	Classificação e previsão de risco de inadimplência ou fraude	Alta eficiência em problemas com múltiplas dimensões	Alto custo computacional e difícil interpretação em bases massivas	[2,3]
<b>Recommender Systems (Sistemas de Recomendação)</b>	Personalização de serviços públicos (saúde, educação, assistência)	Ajusta o serviço ao perfil do usuário, ampliando a efetividade	Risco de reprodução de vieses e discriminação algorítmica	[2, 5, 6]

Fonte: Elaborado pelos autores (2026) com base nas referências do próprio quadro.

Ref. = Referências; [1] = Coeckelbergh (2020); [2] = Grus (2016); [3] = Hastie, Tibshirani e Friedman (2009); [4] = Santos (2024); [5] = Jannuzzi (2024b); [6] = Monteiro e Marrafon (2024).

Nos modelos apresentados no Quadro 1, técnicas estatísticas e de IA demonstram capacidade de identificar riscos, antecipar demandas e segmentar populações, ampliando a eficiência e o potencial prospectivo da ação pública. Destacam-se, em particular, os modelos de

florestas aleatórias, preferíveis às árvores de decisão por combinarem múltiplas árvores simultaneamente, reduzindo o sobreajuste e aumentando a precisão das previsões (Hastie; Tibshirani; Friedman, 2009).

Apesar desses benefícios, a ampliação da capacidade prospectiva produz efeitos ambivalentes: ao fortalecer o planejamento governamental, também pode aumentar a opacidade decisória e favorecer processos de automatização acrítica (Coeckelbergh, 2020; Ellul, 1980). O risco torna-se particularmente relevante quando decisões baseadas em correlações estatísticas são naturalizadas na ação governamental, resultando em modelos de “caixa-preta”, nos quais os critérios decisórios permanecem pouco transparentes ao escrutínio público (Coeckelbergh, 2020; Monteiro; Marrafon, 2024).

Além da modelagem estatística e do uso de IA, a produção de evidências depende de processos rigorosos de organização e padronização informacional. A utilização sistemática de metadados, mantidos em catálogos ou dicionários de dados, garante interoperabilidade, rastreabilidade e consistência semântica entre diferentes bases (Beaulieu, 2010; Date; Kannan; Swamynathan, 2006). Técnicas de *record linkage* (vinculação de registros) e o uso de chaves de integração — como identificadores individuais ou territoriais — permitem conectar registros heterogêneos e construir perfis analíticos consistentes, ampliando o potencial das bases administrativas para análises preditivas e avaliação de Políticas Públicas (Jannuzzi, 2013, 2024a).

Dessa forma, a estrutura de dados estatal combina rotinas de ETL, modelagem estatística e integração de registros para transformar dados brutos em evidências estruturadas (Grus, 2016; Jannuzzi, 2013). O desafio central, contudo, não reside apenas na aplicação de técnicas avançadas, mas na governança dessa capacidade analítica, de modo compatível com fundamentos normativos. Trata-se, portanto, de assegurar que o conhecimento produzido sustente decisões públicas legítimas, auditáveis e institucionalmente responsáveis (Date; Kannan; Swamynathan, 2006; Jannuzzi, 2013, 2018).

## **RISCOS ESTRUTURAIS DA RACIONALIDADE ALGORÍTMICA**

A expansão da base analítica do Estado não gera apenas riscos operacionais, mas também riscos estruturais ligados à própria lógica da técnica. A crítica clássica à autonomia da técnica fornece base teórica para compreender o deslocamento em que ela deixa de ser instrumento subordinado e passa a operar como sistema orientado pela maximização da eficiência como valor intrínseco (Bekkers; Duivenboden; Thaens, 2006; Ellul, 1980; Vanderburg, 2011).

Nesse contexto, a eficiência pode tornar-se critério autossuficiente de validação, independentemente de finalidades normativas, criando o risco de que métricas e modelos passem a determinar o que deve ser decidido. O debate político tende, assim, a deslocar-se para indicadores frequentemente apresentados como neutros (Bekkers; Duivenboden; Thaens, 2006; Jannuzzi, 2018), enquanto se manifesta o fenômeno de dessimbolização cultural descrito por Vanderburg (2011), em que indivíduos perdem referência simbólica e contextual frente à lógica técnica.

Estudos sobre viés algorítmico indicam que modelos treinados em bases históricas tendem a reproduzir padrões de exclusão já presentes nos dados. A aprendizagem automatizada não cria vieses de forma autônoma, mas os reorganiza e amplifica sob a aparência de objetividade estatística. Dessa forma, quando esses sistemas orientam decisões relacionadas a direitos, benefícios ou acesso a Políticas Públicas, seus efeitos tornam-se distributivos e podem comprometer a transparência e a equidade das decisões estatais (Coeckelbergh, 2020, 2021). Em ambientes de governança orientados por dados, tal dinâmica pode obscurecer a dimensão política das decisões públicas sob a aparência de neutralidade técnica (Porter, 1996 *apud* Jannuzzi, 2018).

Essa condição evidencia que o problema não se limita ao funcionamento interno dos algoritmos, mas envolve também os mecanismos institucionais de controle e responsabilização que orientam seu uso. Nesse cenário, a *accountability* algorítmica vai além da exigência de transparência formal. Tornar o código público não garante compreensão nem revisão efetiva, sobretudo em contextos institucionais com letramento algorítmico limitado. A legitimidade das decisões automatizadas depende, portanto, de mecanismos capazes de rastrear processos decisórios, atribuir responsabilidades e assegurar supervisão humana qualificada, evitando um pseudo controle social dos sistemas. Sem essas mediações institucionais, a automação tende a transferir a responsabilidade da esfera política para sistemas técnicos de difícil escrutínio.

Essa dinâmica torna-se ainda mais evidente no campo da avaliação de Políticas Públicas, onde a centralidade dos métodos quantitativos pode conferir autoridade normativa excessiva a indicadores e desenhos experimentais, restringindo o processo decisório, como já mencionado. Nesses casos, o método deixa de informar a decisão e passa a constituir seu principal parâmetro de validação (Jannuzzi, 2018). A racionalidade algorítmica pode ser compreendida, assim, como a convergência institucional de três processos principais: i) Autonomização da eficiência como princípio regulador (Bekkers; Duivenboden; Thaens, 2006); ii) Fragmentação da responsabilidade associada à opacidade técnica (Coeckelbergh,

2020); iii) Normatividade implícita do método, pela qual números e métricas adquirem autoridade em contextos sem mandato político direto (Porter, 1996 *apud* Jannuzzi, 2018).

Nesse contexto, o “quantificacionismo”, discutido por Jannuzzi (2018) a partir de Porter (1996), descreve o processo de transformar decisões públicas em métricas padronizadas, usando a chamada “objetividade mecânica” para conferir autoridade técnica a burocratas sem mandato político direto. Essa centralidade dos indicadores cria uma aparência de neutralidade, que tende a ocultar escolhas normativas, valores e interesses políticos na definição de variáveis, critérios e metas, e ao traduzir problemas complexos em medidas comparáveis — voltadas à legibilidade social necessária à ação estatal — a análise acaba reduzindo fenômenos sociais a dimensões estritamente mensuráveis, marginalizando contextos históricos e desigualdades estruturais que métricas, sob perspectivas reducionistas e positivistas, frequentemente ignoram.

O risco estrutural reside, portanto, na possibilidade de a infraestrutura analítica operar como instância decisória relativamente autônoma. Quando decisões públicas passam a ser apresentadas como consequências técnicas inevitáveis, a técnica deixa de funcionar como instrumento da deliberação democrática e passa a delimitar o próprio espaço do possível. Esses riscos afetam diretamente a democracia digital, pois a concentração da capacidade interpretativa em sistemas analíticos tende a reduzir o espaço para deliberação pública e participação cidadã.

Por fim, a sofisticação estatística não elimina a dimensão interpretativa da decisão pública; ao contrário, exige maior reflexividade institucional para que correlações estatísticas não sejam tratadas como fundamentos normativos absolutos (Jannuzzi, 2018; Coeckelbergh, 2020). Nesse cenário, gestores, pesquisadores, formuladores de políticas e tomadores de decisão — todos aqueles cujo trabalho impacta o coletivo — precisam desenvolver maior consciência crítica para que a técnica seja utilizada em favor das pessoas, evitando transformá-las em meras peças na ‘engrenagem’ de um sistema autônomo, na perspectiva elluliana.

## **EVIDÊNCIAS DA RACIONALIDADE ALGORÍTMICA**

Três casos amplamente documentados na literatura sobre Governança Algorítmica ilustram de forma clara os riscos institucionais associados à decisão pública orientada por modelos preditivos.

Em 2014, uma adolescente negra da Flórida foi presa por tentar levar uma bicicleta e um patinete de criança, enquanto um adulto branco com histórico de roubos armados cometeu furto de valor semelhante. O algoritmo COMPAS (Correctional Offender Management

Profiling for Alternative Sanctions), utilizado para estimar risco de reincidência, classificou a jovem como “alto risco”, enquanto o outro indivíduo foi classificado como “baixo risco”. Dois anos depois, a primeira não reincidiu, enquanto o segundo cometeu novo crime grave. O episódio ilustra como sistemas preditivos, mesmo apresentados como objetivos, podem reproduzir vieses históricos e raciais, transformando desigualdades estruturais em critérios prospectivos. A falta de transparência evidencia o deslocamento da responsabilidade decisória para o plano técnico, reforçando a necessidade de supervisão humana e enquadramentos normativos (Angwin *et al.*, 2016).

Um segundo exemplo ocorreu na Holanda com o sistema SyRI (System Risk Indication), criado para identificar fraudes em benefícios sociais por meio da integração de dados fiscais, trabalhistas, habitacionais e previdenciários. Em 2020, o Tribunal Distrital de Haia declarou seu uso ilegal por violar a privacidade, devido à intrusão informacional e à opacidade do algoritmo (Privacy International, 2020). Assim como no caso anterior, sistemas algorítmicos de detecção de risco podem afetar cidadãos vulneráveis sem mecanismos efetivos de contestação, evidenciando que eficiência e automação devem permanecer subordinadas a valores públicos, supervisão humana e proteção de direitos fundamentais.

Por fim, na Dinamarca, a vigilância algorítmica sobre beneficiários transformou o Estado de bem-estar em um “experimento automatizado”. A unidade de mineração de dados da Administração de Benefícios Públicos integra múltiplos bancos de dados — fiscais, habitacionais, de mobilidade e relações familiares — para identificar possíveis fraudes. Embora o sistema apenas sinalize suspeitas, ele produz perfis detalhados por nacionalidade, endereço e histórico familiar, gerando formas de supervisão altamente intrusivas. O caso reforça que a racionalidade algorítmica, quando politicamente instrumentalizada e opaca, pode deslocar responsabilidades para sistemas técnicos e transformar eficiência em critério normativo, evidenciando a necessidade de salvaguardas institucionais que preservem direitos e legitimidade democráticas (Geiger, 2023).

Esses casos mostram que a ampliação da capacidade analítica estatal pode gerar ganhos de eficiência, mas também riscos de desigualdade e opacidade decisória. A racionalidade algorítmica exige, dessa forma, supervisão normativa e transparência institucional. Portanto, estruturas de governança devem assegurar que a técnica permaneça subordinada a princípios institucionais, conforme o modelo PPVE.

## GOVERNANÇA DIGITAL COMO MEDIAÇÃO INSTITUCIONAL

Após delinear a infraestrutura analítica e seus riscos estruturais, torna-se necessário examinar o papel da Governança Digital como mediação institucional entre técnica e legitimidade democrática. Se a racionalidade algorítmica amplia a capacidade analítica do Estado e, simultaneamente, introduz desafios estruturais, a Governança Digital configura-se como mecanismo de articulação entre poder técnico e legitimidade democrática. Ela não se limita à organização de fluxos informacionais, mas envolve a definição de regras, responsabilidades e limites que enquadram o uso da base de dados.

A Governança Algorítmica assegura rastreabilidade, integridade e controle de acesso, traduzindo exigências técnicas em garantias institucionais, especialmente à luz do *Privacy by Design* (Privacidade desde a concepção) de Cavoukian (1997), que estabelece sete princípios a serem incorporados desde a concepção do sistema: proatividade; privacidade como padrão; incorporação da privacidade ao design; funcionalidade plena; segurança de ponta a ponta; visibilidade e transparência; e respeito ao usuário.

O Quadro 2 diferencia níveis de governança, riscos e funções normativas, evidenciando como a Governança Digital integra proteção informacional e controle algorítmico sob precedência normativa orientada por valores republicanos. Princípios internacionais, como os da United Nations Educational, Scientific and Cultural Organization (UNESCO) (2018), reforçam a centralidade de direitos humanos, supervisão significativa e responsabilização institucional. Esses níveis de governança são sistematizados no quadro a seguir, que explicita seus objetos, riscos e funções normativas.

Quadro 2 – Níveis de governança na Administração Pública orientada por dados.

Dimensão	Governança de Dados	Governança Algorítmica	Governança Digital (Integradora)	Ref.
<b>Objeto central</b>	Bases de dados e fluxos informacionais	Sistemas automatizados e modelos decisórios	Arquitetura institucional do uso da técnica	[1-4]
<b>Foco regulatório</b>	Coleta, tratamento, armazenamento e compartilhamento	Explicabilidade, auditabilidade e supervisão humana	Integração entre valores públicos, regras e técnica	[1, 5, 6]
<b>Risco predominante</b>	Uso indevido e assimetria informacional	Opacidade decisória e viés algorítmico	Autonomização sistêmica da técnica	[2, 7]
<b>Instrumentos típicos</b>	LGPD, controle de acesso, anonimização, metadados	Revisão humana, avaliação de impacto, <i>accountability</i>	Hierarquização normativa e mediação institucional	[1, 8, 9]
<b>Nível de atuação</b>	Infraestrutura informacional	Processo decisório automatizado	Sistema institucional como um todo	[3, 8, 9]
<b>Função normativa</b>	Proteção de direitos informacionais	Contenção da decisão automatizada	Subordinação da técnica a valores republicanos	[7]

**Fonte:** Elaborado pelos autores (2026) com base nas referências do próprio quadro.

Ref = Referências; [1] = Brasil (2018); [2] = Coeckelbergh (2020); [3] = Date, Kannan e Swamynathan (2006); [4] = Dunleavy e Margetts (2023); [5] = Jannuzzi (2024a); [6] = Jannuzzi (2024b); [7] = Jannuzzi (2018); [8] = Jannuzzi (2013); [9] = Jannuzzi (2016)

A análise do Quadro 2 evidencia que a Governança de Dados institui direitos, transparência e responsabilização, garantindo que a técnica permaneça subordinada a normas e à supervisão humana. Em decisões automatizadas, exige-se explicabilidade, auditabilidade e revisão contínua, prevenindo que algoritmos operem de forma autônoma e reforçando o juízo político na Administração Pública digital.

## **CLASSIFICAÇÃO INTEGRADA DE VULNERABILIDADE SOCIAL: UM CASO HIPOTÉTICO**

A referência a valores republicanos baseia-se na Constituição da República Federativa do Brasil de 1988 (Brasil, 1988), que define fundamentos e objetivos orientadores da atuação estatal, como dignidade humana, cidadania, promoção do bem comum e redução de desigualdades (arts. 1º e 3º). Princípios administrativos — legalidade, impessoalidade, moralidade, publicidade e eficiência (art. 37) — delimitam o exercício do poder público. No contexto contemporâneo de digitalização da Administração Pública, tais princípios também orientam o desenvolvimento e o uso de infraestruturas informacionais e sistemas analíticos empregados na gestão estatal. Sistemas automatizados devem, portanto, operar em conformidade com esses parâmetros, promovendo direitos e o interesse público.

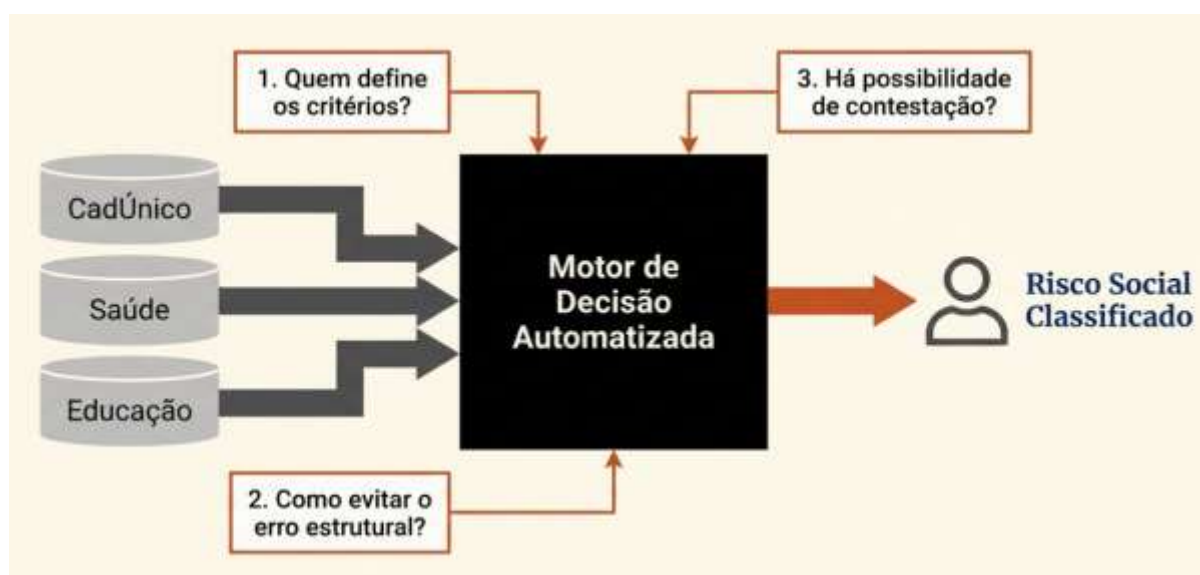
O modelo PPVE manifesta-se em contextos de governança orientada por dados, evidenciando três dimensões principais: valorativa (política) (priorização de fundamentos constitucionais), cognitiva (reconhecimento da complexidade territorial das políticas) e epistêmica (pluralidade de evidências) (Jannuzzi, 2024a, 2024b). O exemplo a seguir funciona como exercício analítico, demonstrando que a legitimidade da infraestrutura depende da precedência normativa que orienta sua utilização.

Considere a hipótese de implementação de um Sistema Integrado de Classificação de Vulnerabilidade Social, destinado a priorizar famílias para Políticas Públicas distributivas. O arranjo articula diferentes bases administrativas — CadÚnico, registros de atenção primária em saúde, frequência escolar e indicadores territoriais — por meio de integração de dados e *record linkage* probabilístico. A consolidação dessas informações permite construir perfis multidimensionais de vulnerabilidade e alimentar um modelo preditivo responsável por classificar o risco social de cada unidade familiar. A arquitetura técnica envolve três etapas

principais: 1) Extração, padronização e integração dos dados via *pipelines* (fluxos de processamento) de ETL; 2) Vinculação de registros heterogêneos para identificar correspondências entre indivíduos, domicílios e territórios; 3) Modelagem analítica destinada a estimar a probabilidade de privação futura ou agravamento da vulnerabilidade.

O resultado é um escore sintético que orienta a priorização de acesso a benefícios, serviços ou acompanhamento intensivo. A Imagem 2 ilustra de forma esquemática essa arquitetura hipotética.

Imagem 2 – Arquitetura hipotética de classificação integrada de vulnerabilidade social.



**Fonte:** Elaborado pelos autores (2026), com auxílio da ferramenta de Inteligência Artificial generativa Nano Banana 2 (Google). **Nota:** O diagrama integra três bases administrativas (CadÚnico, Saúde e Educação) usando record linkage probabilístico e modelagem analítica para gerar um escore de risco social, orientando a priorização de Políticas Públicas.

A análise da Imagem 2 evidencia a ampliação da capacidade diagnóstica do Estado. A integração de dados dispersos supera a fragmentação administrativa, identifica padrões invisíveis à análise isolada e aprimora a focalização de Políticas Públicas. Contudo, o modelo apresenta fragilidades inerentes à classificação automatizada, que se articulam em três questões centrais, baseadas em Jannuzzi (2018, 2024a, 2024b) e Coeckelbergh (2020): i) Definição de critérios – vulnerabilidade é um conceito normativo antes de ser estatístico. A seleção de variáveis e a definição de limiares refletem escolhas valorativas que precedem a modelagem; sem a explicitação desses fundamentos, a técnica tende a naturalizar opções políticas sob a aparência de neutralidade matemática ou “objetividade mecânica” (Porter, 1996 *apud* Jannuzzi, 2018); ii) Erro estrutural – falhas em processos de *record linkage* e na limpeza de dados podem gerar exclusões invisíveis e corroer a validade das evidências (Grus, 2016). Além disso,

modelos treinados com bases históricas assimétricas tendem a reproduzir e consolidar desigualdades sob forma algorítmica (Coeckelbergh, 2020); iii) Contestabilidade – a atribuição de escores de risco afeta o acesso a direitos, exigindo mecanismos de revisão humana qualificada e explicabilidade algorítmica para que a decisão seja inteligível e passível de impugnação (Coeckelbergh, 2020; Jannuzzi, 2024a, 2024b).

O caso hipotético ilustra como infraestruturas analíticas baseadas em dados podem ampliar a capacidade diagnóstica do Estado. Ao integrar registros administrativos dispersos, produzir classificações analíticas e converter essas classificações em critérios de priorização de Políticas Públicas, tais sistemas passam a estruturar os modos pelos quais o Estado conhece e interpreta fenômenos sociais complexos, como a vulnerabilidade.

## DISCUSSÃO

A Ciência de Dados pode ser compreendida como componente de uma infraestrutura epistêmica estatal, na medida em que organiza institucionalmente a produção, a interpretação e a circulação de evidências que estruturam o diagnóstico e a intervenção em problemas públicos. Nesse sentido, sistemas de dados, bases informacionais e capacidades analíticas passam a compor um arranjo institucional que amplia a capacidade do Estado de interpretar a realidade social e orientar a ação governamental com base em evidências.

Nessa perspectiva, a precedência normativa, a supervisão institucional e princípios constitucionais — como equidade, dignidade humana e proteção de direitos — configuram condições estruturantes para o uso legítimo da infraestrutura analítica, ao subordinar a técnica a fundamentos normativos explicitamente definidos (Jannuzzi, 2024a, 2024b). Tal condição relaciona-se à necessidade de mediação institucional e de salvaguardas democráticas na Governança Algorítmica, especialmente em contextos de crescente automatização das decisões públicas (Monteiro; Marrafon, 2024).

Os exemplos do COMPAS, SyRI e da experiência dinamarquesa evidenciam riscos associados à racionalidade algorítmica, como opacidade decisória, deslocamento do juízo político e reprodução de desigualdades históricas (Angwin *et al.*, 2016; Coeckelbergh, 2020). Esses casos demonstram que sistemas apresentados como tecnicamente neutros podem naturalizar padrões sociais desiguais, transferindo responsabilidade política para decisões técnicas de difícil contestação. Embora modelos normativos como o PPVE proponham salvaguardas institucionais, sua implementação enfrenta tensões persistentes entre eficiência, automação e valores republicanos, frequentemente agravadas por limitações de supervisão institucional (Srnicek, 2017; Zuboff, 2019).

A prática revela ainda lacunas sociotécnicas relevantes. Enquanto decisões automatizadas podem permanecer pouco compreensíveis, a revisão humana tende a ser fragmentada e a contestabilidade depende de capacidades técnicas frequentemente restritas. A dependência de dados históricos e de modelos algorítmicos complexos mantém riscos que a estrutura normativa, por si só, não elimina (Selbst *et al.*, 2019).

Sob a perspectiva da teoria democrática, a precedência normativa busca preservar a legitimidade das decisões públicas e evitar que a racionalidade algorítmica restrinja o espaço deliberativo ou amplie assimetrias epistêmicas na tomada de decisão (Habermas, 1992; Estlund, 2008; Landemore, 2013). Ainda assim, a crescente concentração global de infraestrutura de dados e capacidade computacional limita a soberania epistêmica do Estado, indicando que mesmo arranjos normativos consistentes podem revelar-se insuficientes diante de desafios técnicos e estruturais externos (Srnicek, 2017; Zuboff, 2019).

## LIMITAÇÃO E PESQUISAS FUTURAS

Como limitação do presente estudo, destaca-se a sua natureza teórico-analítica, que não envolve investigação empírica sistemática. A proposta da infraestrutura epistêmica estatal como categoria analítica abre, contudo, possibilidades relevantes de investigação empírica sobre a relação entre dados, governança e capacidade estatal. Pesquisas futuras podem explorar, por exemplo: i) de que maneira diferentes arranjos institucionais estruturam a produção e o uso de evidências em Políticas Públicas, especialmente no que se refere à priorização de valores republicanos e objetivos normativos; ii) como diferentes arquiteturas de infraestrutura epistêmica estatal — incluindo integração de bases administrativas, *pipelines* analíticos e sistemas de classificação algorítmica — influenciam a forma como evidências são produzidas e estabilizadas institucionalmente na Administração Pública; e iii) em que medida a configuração da infraestrutura epistêmica estatal afeta a capacidade do Estado de antecipar problemas públicos e estruturar diagnósticos prospectivos na formulação de Políticas Públicas.

## CONSIDERAÇÕES FINAIS

Compreender a Ciência de Dados como infraestrutura epistêmica do Estado permite deslocar o debate da mera eficiência técnica para a legitimidade das decisões públicas orientadas por evidências. Nesse enquadramento, a capacidade analítica estatal não se limita à aplicação de ferramentas estatísticas ou algoritmos preditivos, mas envolve a construção institucional de arquiteturas informacionais capazes de produzir diagnósticos socialmente relevantes e sustentar a formulação, implementação e avaliação de Políticas Públicas.

A articulação entre valores públicos, governança institucional e infraestrutura analítica evidencia que a produção de evidências não constitui um processo neutro, mas um arranjo sociotécnico cuja legitimidade depende da precedência normativa que orienta seu uso na ação estatal. Ao organizar a produção, integração e interpretação de dados em larga escala, essas infraestruturas passam a estruturar os próprios modos pelos quais o Estado conhece, classifica e intervém em problemas públicos.

Essa transformação produz efeitos ambivalentes. Se, por um lado, amplia a capacidade diagnóstica, o planejamento e a focalização de Políticas Públicas, por outro introduz desafios relacionados à opacidade decisória, à concentração de infraestrutura informacional e à limitação da contestabilidade pública. Nesses contextos, preservar a precedência normativa dos valores republicanos torna-se condição essencial para que a infraestrutura analítica funcione como instrumento de fortalecimento da capacidade estatal, e não como mecanismo de deslocamento do juízo político para a lógica autônoma da técnica.

Assim, a consolidação da Ciência de Dados como infraestrutura epistêmica do Estado depende da capacidade institucional de integrar competência analítica, governança democrática e responsabilidade pública na produção e no uso de evidências, garantindo que a expansão da capacidade informacional estatal permaneça orientada por princípios republicanos e compromissos democráticos de transparência, equidade e justiça social.

## REFERÊNCIAS

ALLEN, Danielle S. **Justice by means of democracy**. Chicago; London: The University of Chicago Press, 2023.

ANGWIN, Julia *et al.* **Machine bias**: there's software used across the country to predict future criminals. And it's biased against blacks. ProPublica, 23 maio 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 5 mar. 2026.

BEAULIEU, Alan. **Aprendendo SQL**: dominando os fundamentos de SQL. 2. ed. Tradução: Edgard B. Damiani. Revisão gramatical: Lia Gabriele Regius. São Paulo: Novatec, 2010.

BEKKERS, Victor; DUIVENBODEN, Hein van; THAENS, Marcel (eds.). **Information and communication technology and public innovation**: assessing the ICT-driven modernization of public administration. Amsterdam: IOS Press, 2006.

BRASIL. Presidência da República. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/constituicao/constituicaocompilado.htm](https://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm). Acesso em: 06 mar. 2026.

BRASIL. Presidência da República. **Lei nº 12.527, de 18 de novembro de 2011**. Regula o acesso a informações. Brasília, DF: Presidência da República, 2011. Disponível em: <https://www.gov.br/saude/pt-br/acesso-a-informacao/banco-de-precos/legislacao/lei-no-12-527-de-18-de-novembro-de-2011.pdf/view>. Acesso em: 6 mar. 2026.

BRASIL. Presidência da República. **Lei nº 13.709, de 14 de agosto de 2018**. Lei geral de proteção de dados pessoais (LGPD). Diário Oficial da União: Brasília, DF, 2018. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm). Acesso em: 6 mar. 2026.

CAVOUKIAN, Ann. **Privacy by design: the 7 foundational principles – implementation and mapping of fair information practices**. International Association of Privacy Professionals, 1997. Disponível em: [https://prod.iapp.org/media/pdf/resource\\_center/pbd\\_implement\\_7found\\_principles.pdf](https://prod.iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf). Acesso em: 6 mar. 2026.

CINGOLANI, Luciana. **The state of state capacity: a review of concepts, evidence, and measures**. Maastricht: Maastricht Graduate School of Governance, 2013. (UNU-MERIT Working Paper Series, #2013-053). Disponível em: <https://unu-merit.nl/publications/wppdf/2013/wp2013-053.pdf>. Acesso em: 6 mar. 2026.

COECKELBERGH, Mark. **AI ethics**. Cambridge: MIT Press, 2020.

COECKELBERGH, Mark. Time machines: artificial intelligence, process, and narrative. **Philosophy & Technology**, v. 34, n. 4, p. 1623–1638, 2021.

DATE, C. J.; KANNAN, A.; SWAMYNATHAN, S. **An introduction to database systems**. 8th ed. Boston: Pearson Education, 2006. Disponível em: <https://pearsoned.co.in/prc/book/c-j-date-an-introduction-database-systems-8e-8/9788177585568>. Acesso em: 12 mar. 2026.

DUNLEAVY, Patrick; MARGETTS, Helen. Data science, artificial intelligence and the third wave of digital era governance. **Public Policy and Administration**, [s. l.], v. 40, n. 2, p. 185–214, 2023.

ESTLUND, David. **Democratic authority: a philosophical framework**. Princeton: Princeton University Press, 2008.

ELLUL, Jacques. **The technological system**. Tradução: J. Neugroschel. New York: The Continuum Publishing Corporation, 1980.

FELTRIN, Fernando. **Análise exploratória de dados com Python, Pandas e Numpy**. 2021. Ebook. Disponível em: <https://www.amazon.com.br/An%C3%A1lise-Explorat%C3%B3ria-Dados-Python-Pandas-ebook/dp/B098DF8BZ5>. Acesso em: 10 mar. 2026.

FUKUYAMA, Francis. What is governance?. **Governance**, [s. l.], v. 26, n. 3, p. 347–368, 2013.

GEIGER, Gabriel. **How Denmark’s welfare state became a surveillance nightmare**. WIRED, 7 mar. 2023. Disponível em: <https://www.wired.com/story/algorithms-welfare-state-politics/>. Acesso em: 5 mar. 2026.

GRUS, Joel. **Data Science do Zero**. Tradução: Starlin Alta Editora e Consultoria Eireli. São Paulo: Alta Books, 2016.

HABERMAS, Jürgen. **Between facts and norms: contributions to a discourse theory of law and democracy**. Cambridge: MIT Press, 1992.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome H. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. New York, NY: Springer, 2009.

JANNUZZI, Paulo de Martino. **Indicadores sociais no Brasil: conceitos, fontes de dados e aplicações para formulação e avaliação de políticas públicas e elaboração de estudos socioeconômicos**. 3. ed., 2. impressão. Rio de Janeiro: Alínea, 2009.

JANNUZZI, Paulo de Martino. **Monitoramento e avaliação de programas sociais no Brasil e América Latina: apontamentos conceituais, considerações metodológicas e reflexões sobre as práticas**. Estudo Técnico n.º 07/2013. Brasília: Ministério do Desenvolvimento Social e Combate à Fome, Secretaria de Avaliação e Gestão da Informação, 2013. Disponível em: <https://aplicacoes.mds.gov.br/sagicad/pesquisas/documentos/estudo/41.pdf>. Acesso em: 6 mar. 2026.

JANNUZZI, Paulo de Martino. Eficiência econômica, eficácia procedural ou efetividade social: três valores em disputa na Avaliação de Políticas e Programas Sociais. **Desenvolvimento em Debate**, v. 4, n. 1, p. 117-142, 2016.

JANNUZZI, Paulo de Martino. Mitos do desenho quasi-experimental na avaliação de programas. **Revista NAU Social**, v. 9, n. 16, p. 76-90, 2018.

JANNUZZI, Paulo de Martino. **Políticas públicas, valores e evidências em tempos de inteligência artificial**. Rio de Janeiro: Editora Alínea, 2024a.

JANNUZZI, Paulo de Martino. **Projeto inteligência artificial aplicada às políticas públicas (IAPP)**. Material técnico. Apresentação. CIAP, 2024b.

JANNUZZI, Paulo de Martino. **Mesa 2 – inteligência artificial e políticas públicas**. Material técnico. Apresentação, CIAP, 2025.

LANDEMORE, Hélène. **Democratic reason: politics, collective intelligence, and the rule of the many**. Princeton; Oxford: Princeton University Press, 2013.

MANN, Michael. The autonomous power of the state: its origins, mechanisms and results. **European Journal of Sociology / Archives Européennes de Sociologie**, v. 25, n. 2, p. 185-213, 1984.

MARK, David; MORISON, John. Governing AI decision-making: balancing innovation and accountability. **Politics and Governance**, v. 13, art. 10245, 2025.

MONTEIRO, Juliana Iunes; MARRAFON, Marco Aurélio. Legitimidade democrática na governança algorítmica. **Revista Direitos Fundamentais & Democracia**, Curitiba, v. 29, n. 1, p. 5-50, 2024.

NOVECK, Beth Simone. **Smart citizens, smarter state**: the technologies of expertise and the future of governing. Cambridge, MA; London: The Belknap Press of Harvard University Press, 2015.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **OECD guidelines on the protection of privacy and transborder flows of personal data**. Paris: OECD Publications Service, 2001. Disponível em: [https://www.oecd.org/en/publications/2002/02/oecd-guidelines-on-the-protection-of-privacy-and-transborder-flows-of-personal-data\\_g1gh255f.html](https://www.oecd.org/en/publications/2002/02/oecd-guidelines-on-the-protection-of-privacy-and-transborder-flows-of-personal-data_g1gh255f.html). Acesso em: 6 mar. 2026.

PRIVACY INTERNATIONAL. **The SyRI case**: a landmark ruling for benefits claimants around the world. Privacy International, 24 fev. 2020. Disponível em: <https://privacyinternational.org/news-analysis/3363/syri-case-landmark-ruling-benefits-claimants-around-world>. Acesso em: 5 mar. 2026.

SACKETT, David L. *et al.* Evidence based medicine: what it is and what it isn't. **BMJ**, London, v. 312, n. 7023, p. 71–72, 1996.

SANTOS, Tibério Celso Gomes dos. **A governança de dados no contexto do Estado Digital e a tomada de decisão baseada em evidências**: a revolução dos dados na administração tributária. 2024. 311 f. Tese (Doutorado em Direito) – Faculdade de Direito, Universidade Federal do Ceará, Fortaleza, 2024. Disponível em: <https://repositorio.ufc.br/ri/handle/riufc/79564>. Acesso em: 22 fev. 2026.

SELBST, Andrew D. *et al.* Fairness and abstraction in sociotechnical systems. **Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency**, Atlanta, GA, January 29–31, 2019. ACM, 2019. Disponível em: <https://doi.org/10.1145/3287560.3287598>. Acesso em: 02 mar. 2026.

SRNICEK, Nick. **Platform capitalism**. Cambridge, UK; Malden, MA: Polity Press, 2017.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (UNESCO). **Principles on personal data protection and privacy**. Disponível em: <https://www.unsystem.org/privacy-principles>. Acesso em: 22 fev. 2026.

VAN DERBURG, Willem H. **Our war on ourselves**: rethinking science, technology, and economic growth. Toronto; Buffalo; London: University of Toronto Press, 2011.

ZUBOFF, Shoshana. **The age of surveillance capitalism**: the fight for a human future at the new frontier of power. New York: PublicAffairs, 2019.

WICKHAM, Hadley; ÇETINKAYA-RUNDEL, Mine; GROLEMUND, Garrett. **R for data science**: visualize, model, transform, tidy, and import data. 2. ed. Sebastopol, CA: O'Reilly Media, 2023. E-book. Disponível em: <https://www.oreilly.com/library/view/r-for-data/9781492097396/>. Acesso em: 11 mar. 2026.