Casi 17ª Edição
Congresso de Administração, Sociedade e Inovação

Artigo tecnológico
De 03 a 05 de dezembro de 2025

ÉTICA E GOVERNANÇA PARA INTERVENÇÕES DIGITAIS NO ESTADO – ÉGIDE: Um Framework *Nudges* Digitais na Administração Pública

Marcelo Tenório Malta
Doutorando em Contabilidade e Administração na FUCAPE
marcelotmalta@outlook.com

Resumo: O uso de *nudge*s digitais personalizados mediados por algoritmos tem se expandido como estratégia relevante para induzir comportamentos em políticas públicas. Apesar do potencial de eficiência, a ausência de protocolos éticos claros, a opacidade algorítmica e as exigências legais da LGPD evidenciam riscos que podem comprometer sua legitimidade. Este artigo apresenta o ÉGIDE – Ética e Governança para Intervenções Digitais no Estado, um framework tecnológico estruturado em cinco pilares: diagnóstico, governança, validação, execução e monitoramento. O diagnóstico identificou lacunas recorrentes em práticas institucionais, como fragmentação de responsabilidades, falta de auditoria e insuficiência de salvaguardas éticas. A proposta organiza instrumentos aplicáveis, incluindo matrizes de governança, protocolos de validação e planos de rastreabilidade, que orientam gestores na implementação responsável de *nudge*s digitais. Como contribuição, o ÉGIDE oferece um modelo replicável que alinha inovação tecnológica à transparência, ao compliance normativo e à confiança pública, favorecendo a institucionalização de práticas éticas em políticas digitais.

Palavras-Chave: *Nudge*s digitais; Governança algorítmica; Ética em IA; LGPD; Administração pública.

1. Introdução

A transformação digital tem redefinido a forma como governos e organizações influenciam comportamentos sociais. Tecnologias baseadas em big data, inteligência artificial (IA) e aprendizado de máquina ampliaram a capacidade de personalização das políticas públicas, permitindo a adoção de *nudge*s digitais como instrumentos de arquitetura de escolhas (Thaler & Sunstein, 2009). Esses "empurrões" suaves vêm sendo aplicados em larga escala para incentivar a adesão a programas governamentais, melhorar a arrecadação tributária e promover comportamentos desejáveis em áreas como saúde, educação e mobilidade (DellaVigna & Linos, 2020; Dolan et al., 2012).

Entretanto, o uso de *nudge*s mediados por algoritmos traz consigo um dilema central. De um lado, estudos apontam ganhos expressivos de eficiência, com aumento da efetividade de políticas e redução de custos operacionais (Hallsworth et al., 2017). De outro, a ausência de protocolos éticos claros, o risco de manipulação e a opacidade algorítmica podem minar a confiança pública e comprometer a legitimidade institucional dessas intervenções (Yeung, 2018; Jobin et al., 2019). Esse paradoxo evidencia que a eficácia técnica, por si só, não garante a sustentabilidade das práticas digitais quando não acompanhada de salvaguardas de transparência, justiça e responsabilização (Floridi et al., 2018).

No Brasil, esse debate assume contornos específicos devido às exigências legais da Lei Geral de Proteção de Dados (Lei nº 13.709/2018) e a propostas de regulação da inteligência artificial em tramitação no Congresso Nacional (Van Ooijen et al., 2019). Além das questões regulatórias, muitos órgãos públicos enfrentam lacunas práticas, como a fragmentação institucional, a falta de mecanismos de auditoria e a ausência de protocolos de comunicação clara com os cidadãos (Morley et al., 2020; OECD/CAF, 2022). Esses fatores reforçam a urgência de frameworks que conciliem inovação tecnológica com legitimidade democrática e segurança jurídica.

Nesse contexto, este artigo apresenta o ÉGIDE – Ética e Governança para Intervenções Digitais no Estado. Trata-se de um produto tecnológico concebido para orientar gestores públicos na aplicação responsável de *nudge*s digitais personalizados. O ÉGIDE estrutura-se em cinco pilares – diagnóstico, governança, validação, execução e monitoramento – oferecendo ferramentas operacionais que traduzem princípios éticos em práticas aplicáveis. Seu propósito é alinhar inovação digital, conformidade regulatória e confiança social, fornecendo um modelo replicável para o setor público brasileiro.

Dessa forma, a contribuição deste trabalho situa-se no enfoque de melhoria e extrapolação: propõe novos protocolos de governança para práticas já conhecidas de nudging e adapta recomendações internacionais ao contexto regulatório e institucional da administração pública no Brasil (Motta, 2017; 2022). O objetivo central é apresentar um arcabouço prático que auxilie gestores na tomada de decisão ética e eficiente, conciliando desempenho tecnológico com legitimidade institucional.

2. Contexto do problema (ou da oportunidade)

A aplicação de *nudge*s digitais personalizados está na interseção de três dimensões críticas: ciência comportamental, governança algorítmica e ética em inteligência artificial. Cada uma dessas frentes oferece fundamentos essenciais para compreender tanto o potencial quanto os riscos envolvidos em intervenções baseadas em dados.

Do ponto de vista da ciência comportamental, reconhece-se que decisões humanas são influenciadas por vieses cognitivos e heurísticas, e que pequenas alterações na arquitetura de escolhas podem induzir comportamentos socialmente desejáveis sem restringir a liberdade individual (Sugden, 2009; DellaVigna & Linos, 2020; Dolan et al., 2012). Estratégias desse tipo já demonstraram resultados relevantes em políticas públicas, como o aumento da arrecadação tributária no Reino Unido por meio de mensagens personalizadas enviadas a contribuintes em atraso (Hallsworth et al., 2017).

A integração de algoritmos amplia o alcance e a eficácia dessas intervenções, permitindo personalização em grande escala. Contudo, levanta preocupações sobre opacidade algorítmica, riscos de manipulação, discriminação e violação da autonomia individual (Yeung, 2018; Jobin et al., 2019). A literatura internacional enfatiza que a ausência de mecanismos claros de explicabilidade e prestação de contas compromete a legitimidade das políticas digitais (Floridi et al., 2018; OECD, 2022).

Esse debate é particularmente relevante no Brasil, onde o arcabouço regulatório impõe desafios adicionais. A Lei Geral de Proteção de Dados (Lei nº 13.709/2018) estabelece que decisões automatizadas relevantes devem ser transparentes e passíveis de revisão. Além disso, discussões legislativas em andamento sobre inteligência artificial (PL nº 21/2020) reforçam a necessidade de salvaguardas institucionais. Apesar disso, muitas organizações públicas ainda carecem de protocolos mínimos para avaliar riscos éticos ou assegurar conformidade regulatória (Van Ooijen et al., 2019).

O ambiente institucional revela lacunas práticas que dificultam a adoção segura de *nudge*s digitais personalizados. Entre elas, destacam-se a fragmentação entre áreas técnicas e jurídicas, a ausência de auditorias específicas para algoritmos, a falta de documentação dos critérios de segmentação e a carência de canais efetivos de contestação pelos cidadãos (Morley et al., 2020; OECD/CAF, 2022). Essas falhas elevam o risco de judicialização, perdas reputacionais e desconfiança social, reduzindo o potencial de inovação digital no setor público.

Assim, o contexto do problema pode ser sintetizado em um paradoxo: enquanto as tecnologias digitais oferecem instrumentos mais poderosos para promover comportamentos desejáveis, sua adoção desestruturada e sem governança pode gerar efeitos contraproducentes. Esse cenário evidencia a urgência de modelos operacionais que articulem eficácia técnica com legitimidade ética, transformando recomendações normativas em práticas aplicáveis no cotidiano da gestão pública.

3. Diagnóstico do problema (ou da oportunidade)

O diagnóstico das práticas de uso de *nudge*s digitais evidencia que os ganhos de eficácia não têm sido acompanhados pela consolidação de estruturas de governança adequadas. Embora estudos internacionais mostrem que a personalização pode elevar em até 33% a efetividade de programas públicos (DellaVigna & Linos, 2020; Capasso & Umbrello, 2022), a aplicação prática revela fragilidades significativas em termos de legitimidade e conformidade regulatória.

A primeira causa identificada é a opacidade algorítmica. Muitos órgãos não dispõem de mecanismos de documentação ou explicabilidade dos critérios utilizados na segmentação de públicos, o que compromete a transparência e dificulta a prestação de contas (Yeung, 2018; Jobin et al., 2019). Sem protocolos claros de auditoria, os riscos de manipulação ou discriminação tendem a se ampliar, sobretudo em intervenções que envolvem dados sensíveis.

A segunda fragilidade está na fragmentação institucional. Em diversos contextos, áreas de tecnologia da informação, jurídico e comunicação atuam de forma isolada, sem protocolos unificados de desenho e monitoramento das intervenções. Essa desarticulação resulta em ausência de matrizes de responsabilidade, registros de critérios de segmentação e rotinas de acompanhamento contínuo, aumentando a vulnerabilidade a falhas e litígios (Morley et al., 2020; OECD, 2022).

O terceiro ponto crítico refere-se à conformidade regulatória insuficiente. Embora a LGPD exija que decisões automatizadas relevantes sejam explicáveis e revisáveis, muitas organizações ainda não implementaram processos que garantam o cumprimento efetivo desses requisitos (Lei nº 13.709/2018; Van Ooijen et al., 2019). Adicionalmente, a ausência de normativos específicos sobre inteligência artificial no Brasil amplia a incerteza regulatória, expondo gestores a riscos jurídicos.

Por fim, observa-se a carência de mecanismos de accountability e comunicação com o cidadão. Frequentemente, não há canais acessíveis para que os indivíduos compreendam os critérios de personalização ou contestem os resultados das intervenções. Essa lacuna reduz a confiança social e aumenta a percepção de manipulação, com potencial para comprometer a efetividade de longo prazo das políticas digitais (OECD/CAF, 2022).

Portanto, o diagnóstico demonstra que, embora tecnicamente promissores, os *nudge*s digitais carecem de uma base institucional robusta que assegure legitimidade ética, proteção de direitos e alinhamento regulatório. A superação dessas fragilidades depende da criação de um arcabouço estruturado de governança, capaz de traduzir princípios normativos em instrumentos práticos para orientar gestores públicos na aplicação responsável das intervenções comportamentais.

4. Proposta da solução do problema (ou do aproveitamento da oportunidade)

Como solução, temos a propositura do ÉGIDE (Ética e Governança para Intervenções Digitais no Estado). O ÉGIDE é um framework tecnológico estruturado em cinco pilares interdependentes: 1) diagnóstico, 2) governança, 3) validação, 4) implementação e 5) transparência (Figura 1). Diferentemente de modelos meramente conceituais, o ÉGIDE oferece instrumentos operacionais que permitem sua adoção prática por gestores públicos, como o *Canvas de Intervenção Comportamental*, o *Behaviour Change Wheel (BCW)*, a *Matriz de Governança Ética*, protocolos de validação e a matriz RACI.

Figura 1. Etapas para aplicação de *Nudges* com Governança Ética – ÉGIDE
Fonte: Elaboração própria

A proposta é que o ÉGIDE funcione como um manual de uso, capaz de traduzir princípios normativos em práticas aplicáveis, permitindo que qualquer órgão, independentemente de seu nível de maturidade digital, possa adotar intervenções comportamentais alinhadas à legislação e aos valores democráticos.

Para ilustrar sua aplicação, utilizaremos como caso hipotético o exemplo do aumento da adimplência tributária, a partir da experiência realizada pelo *Behavioural Insights Team* (BIT) no Reino Unido. Nesse estudo, o BIT realizou o envio de mensagens personalizadas a contribuintes em atraso, aumentaram significativamente a taxa de pagamento voluntário de impostos (Hallsworth et al., 2017). A partir desse caso documentado, simularemos como o ÉGIDE poderia orientar gestores públicos na concepção, validação e implementação de uma política semelhante no Brasil.

4.1 Primeiro Pilar: Diagnóstico e Alinhamento Estratégico

O diagnóstico é o ponto de partida de qualquer intervenção baseada no ÉGIDE. Seu objetivo é identificar claramente o comportamento a ser influenciado, as barreiras existentes e o alinhamento com os objetivos estratégicos da instituição.

Duas ferramentas são centrais nessa etapa (Figura 2): o Canvas de Intervenção Comportamental, que permite mapear de forma visual os fatores que influenciam a decisão do público, e o Behaviour Change Wheel (BCW), que organiza esses fatores em três dimensões centrais – capacidade, oportunidade e motivação (Michie et al., 2011).

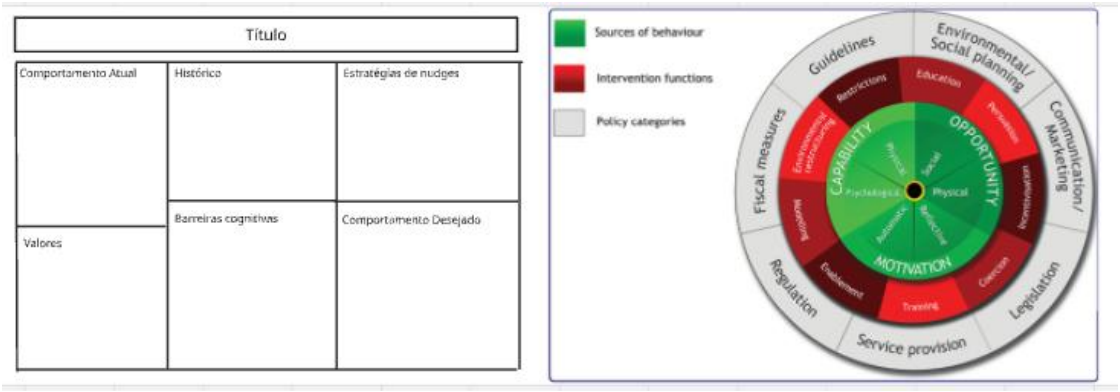

Figura 2 – Exemplos de Canvas (esquerda) e BCW (direita)
Fonte: Canvas – produção própria. Behaviour Change Wheel (Michie et al., 2011)

O Canvas de Intervenção Comportamental é uma ferramenta visual que auxilia no mapeamento dos fatores que influenciam o comportamento atual e na identificação de pontos de intervenção. Ele permite sistematizar: O comportamento atual e o desejado; As barreiras cognitivas, contextuais ou motivacionais que afetam o público-alvo; e As estratégias de *nudge* que podem ser aplicadas de forma ética e eficaz.

Já o *Behaviour Change Wheel* (Michie et al., 2011) é um modelo teórico que organiza as intervenções comportamentais em torno de três componentes centrais: Capacidade, Oportunidade e Motivação (COM-B), que influenciam diretamente o comportamento. A roda orienta o gestor a: Diagnosticar quais desses três componentes estão limitando o comportamento desejado; Selecionar tipos de intervenção (ex: persuasão, educação, modelagem) mais adequados; e Integrar essas escolhas a políticas públicas ou organizacionais existentes.

O Quadro 1 simula os resultados de utilização do Canvas e do BCW no nosso exemplo.

Quadro 1. Exemplo de aplicação do Canvas e do BWC(Behaviour Change Wheel).

Comportamento-alvo: aumento do pagamento voluntário de impostos em atraso.
Canvas de Intervenção:

- **Comportamento atual**: contribuintes atrasam ou ignoram o pagamento.
- **Comportamento desejado**: pagamento espontâneo dentro do prazo.
- **Barreiras:** inércia decisória, percepção de baixa fiscalização.
- **Estratégias de *nudge***: mensagens que destacam normas sociais (“a maioria das pessoas da sua região já pagou seus impostos”).

BCW (COM-B):

- **Capacidade:** os contribuintes sabiam como pagar.
- **Oportunidade:** acesso fácil aos canais de pagamento.
- **Motivação:** baixa, devido à percepção de impunidade.
- **Intervenção escolhida:** uso de **mensagens sociais** para reforçar a motivação.

Fonte: Interpretação dos dados constantes no artigo de Hallsworth et al. (2017).

O uso do *Canvas* ajuda a visualizar essas dimensões e orientar a formulação da intervenção. Já o BCW permite identificar que a principal lacuna não está na capacidade (os contribuintes sabem como pagar) nem na oportunidade (existem canais disponíveis), mas sim na motivação. Assim, a intervenção deve focar em mensagens que reforcem normas sociais e benefícios coletivos.

Esse diagnóstico inicial é fundamental para evitar abordagens genéricas ou mal direcionadas, garantindo que a solução seja construída a partir de evidências concretas e alinhada aos objetivos estratégicos da organização, como aumento da arrecadação com mínima coercitividade e fortalecimento da confiança social.

4.2 Segundo Pilar: Governança Ética e Algorítmica

O segundo pilar busca garantir que a intervenção seja concebida de forma ética, transparente e responsável, antes mesmo de sua implementação. Para isso, propõe-se o uso da Matriz de Governança Ética (Tabela 1), que funciona como checklist operacional a ser preenchido por um comitê interdisciplinar envolvendo áreas de TI, jurídico, compliance, comunicação e atendimento ao cidadão.

Tabela 1. Matriz Simplificada de Governança Algorítmica

Eixo	Questões Orientadoras
Autonomia	A intervenção respeita a liberdade de escolha? Existe opção clara de recusa?
Transparência	O público sabe que está sendo direcionado? O modelo é explicável?
Justiça e equidade	Há risco de tratamento desigual entre grupos? Os dados são representativos?
Privacidade	Os dados foram coletados com base legal e consentimento válido?
Responsabilização	Quem audita os modelos? Há mecanismo de correção ou contestação disponível?

Fonte: Modelo é inspirado em frameworks como o da OCDE, na High-Level Expert Group da União Europeia (2019) e de Jobin et al.(2019), e pode ser adaptado em uma planilha de checklist operacional

Essa matriz avalia dimensões da autonomia, transparência, justiça e equidade, privacidade e responsabilidade. Ao utilizar a matriz, o gestor deve: Personalizar os eixos: adaptar os princípios conforme os valores institucionais e marcos normativos vigentes (ex: LGPD, código de ética institucional); Responder coletivamente às questões: promover oficinas com representantes das áreas envolvidas para preencher a matriz antes do piloto da intervenção; Identificar lacunas e ações corretivas: onde houver respostas negativas ou incertas, estabelecer planos de mitigação ou protocolos adicionais (ex: revisão jurídica, comunicação clara com o cidadão); e Formalizar e documentar: registrar as respostas em uma planilha ou formulário digital e arquivar junto ao processo da política pública, para fins de auditoria e transparência.

No caso ilustrativo (Quadro 2), a matriz permite identificar que os dados utilizados (inadimplência registrada em sistemas fiscais) têm base legal e não configuram uso abusivo. Contudo, apontam também a necessidade de explicitar ao cidadão o motivo da personalização

da mensagem, garantindo transparência ativa. Além disso, recomenda-se a criação de um comitê de auditoria para supervisionar a intervenção, prevenindo riscos éticos e jurídicos.

Quadro 2. Aplicação da Matriz de Governança Algorítmica

Autonomia: os contribuintes podiam optar por não responder às mensagens.
Transparência: há clareza sobre o valor devido, embora não haja comunicação explícita sobre o uso de normas sociais como estratégia.
Justiça: todos os contribuintes receberam mensagens semelhantes, reduzindo risco de discriminação.
Privacidade: os dados usados são de domínio público (inadimplência registrada), respeitando base legal.
Responsabilização: formalização de um comitê de auditoria e publicação dos resultados.

Fonte: Interpretação dos dados constantes no artigo Hallsworth et al.(2017).

Esse pilar garante que a busca por eficiência arrecadatória não se sobreponha a valores democráticos, antecipando possíveis litígios e fortalecendo a legitimidade institucional.

4.3 Terceiro Pilar: Validação Ética e Testes Comportamentais

Nenhuma intervenção deve ser implementada em larga escala sem antes passar por uma etapa de validação. O ÉGIDE propõe que essa etapa seja realizada por meio de testes A/B, protocolos de consentimento proporcional e documentação rigorosa dos critérios de personalização.

4.3.1 A/B *Testing* e grupos de controle

O uso de testes A/B é uma prática consagrada para verificar a eficácia de *nudges*. A ideia central é dividir o público-alvo em pelo menos dois grupos:Um grupo controle, que não recebe o *nudge* ou recebe uma versão neutra; e Um grupo tratamento, que recebe a intervenção com personalização (por exemplo, linguagem social, dados comparativos, mensagens com emojis, entre outros).

Esse método permite comparar, com validade estatística, se houve diferença significativa no comportamento entre os grupos. A recomendação é utilizar indicadores previamente definidos, como taxas de resposta, conversão ou adesão, e garantir uma amostra suficientemente robusta para evitar vieses.

Além disso, a testagem deve prever monitoramento de efeitos adversos não intencionais, como percepção de manipulação, aumento de desconfiança institucional ou impacto desigual em subgrupos vulneráveis.

4.3.2 Protocolos de consentimento e comunicação transparente

Mesmo quando não exigido legalmente, é recomendada a adoção de protocolos de consentimento informado e comunicação clara. Isso fortalece a confiança institucional e está alinhado a princípios de boa governança algorítmica. O consentimento pode ser operacionalizado por mecanismos simples de opt-out, checkbox, ou termo de ciência, especialmente quando a personalização envolve dados sensíveis. A comunicação deve explicar de forma acessível o propósito da intervenção, os critérios utilizados e as opções disponíveis ao cidadão.

Exemplo prático: um lembrete fiscal via SMS pode conter um link com informações sobre como os dados foram utilizados para personalizar a mensagem.

4.3.3 Documentação dos parâmetros de personalização

Para garantir auditabilidade e responsabilização, é essencial manter um repositório com os seguintes elementos: a) Critérios de segmentação: quais características foram utilizadas (ex: idade, localização, histórico de interação); b) Fontes e bases de dados utilizadas; c) Lógica do algoritmo: fluxogramas, regras de decisão, ponderações ou scripts de classificação aplicados; e d) Versões da mensagem: modelos de texto, variações linguísticas, canais utilizados.

No exemplo do Reino Unido, a validação foi feita comparando três grupos: um controle, que recebeu a mensagem padrão; um grupo que recebeu a mensagem personalizada destacando normas sociais locais; e um grupo que recebeu a versão nacional. Os resultados mostraram que a personalização local teve maior impacto (Hallsworth et al., 2017). O Quadro 3 apresenta os resultados destas análises.

Quadro 3. Etapa de desenho, teste e validação

a) A/B Testing e grupos de controle

- **Definição prévia dos indicadores:** além da taxa de pagamento, incluir métricas de percepção pública (ex.: confiança na Receita, sensação de justiça).
- **Grupo controle:** receber a carta padrão, sem elementos adicionais.
- **Grupo tratamento:** receber a carta personalizada com normas sociais (“a maioria dos cidadãos da sua região já pagou seus impostos”).
- **Monitoramento de efeitos adversos:** além da efetividade, coletar dados sobre potenciais efeitos colaterais, como rejeição da mensagem ou aumento de reclamações junto a canais de atendimento.

b) Protocolos de consentimento e comunicação transparente

- **Transparência ativa:** incluir, na carta, uma nota simples: “Esta mensagem foi adaptada com base em dados de adimplência registrados em seu cadastro”.
- **Canal de esclarecimento:** um link ou QR code direcionando para página do órgão, explicando como os dados foram utilizados e oferecendo opção de contato com a administração tributária.
- **Consentimento proporcional:** embora não seja necessário consentimento explícito para comunicação fiscal obrigatória, pode haver a possibilidade de *opt-out* de mensagens adicionais não vinculadas ao cumprimento legal (ex.: lembretes informativos). Isso pode ser configurado no espaço digital do contribuinte ou com opção dentro do sítio do órgão.

c) Documentação dos parâmetros de personalização

- **Critérios de segmentação**: inadimplência registrada nos sistemas fiscais.
- **Fontes de dados**: cadastros administrativos da Receita.
- **Lógica de personalização**: inserção de frases baseadas em normas sociais.
- **Versões da mensagem**: padrão, personalizada regional, personalizada nacional.

- **Registro dos resultados**: taxas de pagamento por grupo, percepção pública (pesquisas pós-intervenção), incidência de reclamações.

Fonte: Interpretação dos dados constantes no artigo Hallsworth et al.(2017).

Além da análise de eficácia, o ÉGIDE orienta que se avaliem também possíveis efeitos adversos, como aumento de reclamações em canais de atendimento ou percepção de manipulação. Esses dados devem ser documentados e considerados na decisão de ampliar a intervenção.

A validação fortalece a legitimidade institucional ao mostrar que a intervenção é baseada em evidências, reduz riscos de falhas e assegura que a inovação seja adotada de forma responsável.

4.4 Quarto Pilar: Implementação com Supervisão e Rastreabilidade

A implementação exige não apenas a execução da intervenção, mas também mecanismos de prestação de contas e rastreabilidade. O ÉGIDE recomenda três práticas centrais: 1) Supervisão cruzada - equipes diferentes das que elaboraram a intervenção devem ser responsáveis por sua avaliação, evitando conflitos de interesse; Registro contínuo - todas as mensagens enviadas, os públicos alcançados e os resultados observados devem ser registrados em sistemas eletrônicos, com logs detalhados; e Auditorias periódicas - tanto internas (unidades de controle e corregedorias) quanto externas (tribunais de contas, defensorias, conselhos de transparência).

Modelos como *o Algorithmic Impact Assessment* do Canadá e as diretrizes da *Netherlands Court of Audit* demonstram que essas práticas são viáveis e contribuem para a legitimidade institucional. Essas auditorias devem avaliar não apenas o desempenho técnico da intervenção, mas também aspectos como: Justiça distributiva (impacto entre diferentes grupos sociais); Transparência e comunicação; e Proporcionalidade do uso de dados.

A institucionalização dessa etapa fortalece a *accountability* algorítmica e ajuda a proteger a organização contra riscos reputacionais, jurídicos e éticos. O quadro 4 apresenta a simulação da quarta fase do EGIDE de implementação em nosso exemplo.

Quadro 4. Princípios Operacionais e uso do AIA

1 - Princípios Operacionais:
Supervisão cruzada: a equipe de ciência comportamental que desenhou as cartas não é responsável pela avaliação final. Um comitê interdisciplinar, incluindo jurídico e auditoria governamental, revisará o processo.
Registro e rastreabilidade: cada carta enviada é registrada em sistema eletrônico, com metadados (data, versão da mensagem, contribuinte alcançado), permitindo reconstruir a intervenção.
Auditoria: auditorias internas podem verificar proporcionalidade e conformidade legal. Auditorias externas (ex.: Tribunal de Contas ou organismos independentes) avaliam a legitimidade do uso de normas sociais em comunicações fiscais.

2 - Levantamento do Impacto Algorítmico (AIA)

Classificação do risco: impacto médio, pois envolve grande volume de contribuintes, mas não utilizava dados sensíveis além da inadimplência já pública. **Salvaguardas:** criação de um comitê de revisão com representantes da área jurídica, ciência de dados e ouvidoria para validar a mensagem antes do envio. **Prestação de contas:** publicação de um relatório acessível, explicando por que a norma social foi usada, como os dados foram tratados e quais eram as alternativas avaliadas.

Fonte: Interpretação dos dados constantes no artigo Hallsworth et al.(2017).

4.5 Quinto Pilar: Transparência Pública e Aprendizado Contínuo

O último pilar busca consolidar a legitimidade e a sustentabilidade da intervenção ao longo do tempo. Para isso, recomenda-se a institucionalização de três práticas: 1)Escuta ativa: pesquisas de percepção junto aos contribuintes, análise de mídias sociais e grupos focais para captar reações à intervenção; 2) Publicação de relatórios de impacto: relatórios semestrais com indicadores de eficácia (taxa de pagamento), equidade (impacto por região, gênero ou renda), satisfação (nível de compreensão, taxa de reclamações) e segurança algorítmica (erros, vieses, revisões); e 3) Revisão periódica dos modelos: ajustes nas mensagens e algoritmos, prevenindo fadiga cognitiva e garantindo alinhamento a mudanças legais ou sociais.

No caso ilustrativo (Quadro 5), a secretaria poderia publicar relatórios em portal de transparência, explicando como as mensagens foram personalizadas, quais resultados alcançaram e quais ajustes foram feitos. Essa prática fortalece a confiança pública e reduz o risco de judicialização.

Quadro 5. Melhoria contínua

Monitoramento contínuo: além das taxas de pagamento, o governo acompanha percepções de confiança na Receita e número de reclamações. **Canal de feedback:** contribuintes podem avaliar a clareza das mensagens por meio de pesquisa rápida (ex.: QR code no documento). **Transparência ativa:** publicação de relatório anual descrevendo metodologia, impacto arrecadatório, riscos e salvaguardas aplicadas. **Ajustes incrementais:** revisão periódica das mensagens para evitar fadiga do contribuinte ou eventual percepção de manipulação.

Fonte: Interpretação dos dados constantes no artigo Hallsworth et al.(2017).

4.6 Síntese da Proposta

O ÉGIDE diferencia-se por oferecer um roteiro passo a passo, aplicável a gestores que precisam adotar intervenções digitais éticas e eficazes. A partir de um único exemplo – o combate à inadimplência tributária – demonstrou-se como o framework orienta todas as etapas, desde o diagnóstico inicial até a prestação de contas final.

Seu valor reside na capacidade de transformar princípios em prática: o Canvas ajuda a estruturar o diagnóstico, a Matriz Ética orienta a governança, os testes A/B validam a intervenção, os registros e auditorias asseguram rastreabilidade e os relatórios de impacto consolidam a transparência (Tabela 2).

Tabela 2. Estrutura do ÉGIDE: Cinco Pilares para *Nudges* Éticos no Estado

Pilar	Etapa do ÉGIDE	Objetivo central	Instrumentos-chave
1. Diagnóstico	Mapeamento do comportamento-alvo e barreiras	Alinhar intervenção aos objetivos institucionais	Canvas de Intervenção, Behaviour Change Wheel
2. Governança	Construção de salvaguardas éticas	Garantir transparência, justiça e responsabilização	Matriz de Governança Ética
3. Validação	Testes A/B, protocolos de consentimento	Avaliar impacto e prevenir efeitos colaterais	Testes controlados, opt-out, documentação
4. Implementação	Execução supervisionada e rastreável	Assegurar integridade, separação de funções e auditabilidade	Comitê de Ética, registros, logs operacionais
5. Transparência	Feedback, prestação de contas e revisões	Reforçar confiança pública e promover melhoria contínua	Relatórios públicos, pesquisas de percepção

Mais do que uma proposta teórica, o ÉGIDE configura-se como produto tecnológico replicável e adaptável, capaz de ser aplicado em diferentes setores da administração pública, como saúde (ex.: aumento da adesão a campanhas de vacinação), educação (ex.: estímulo à frequência escolar) e mobilidade urbana (ex.: incentivo ao transporte público). Em todos esses casos, sua aplicação reforçaria a confiança social, alinhando inovação digital com princípios democráticos e conformidade regulatória.

5. Plano de ações da mudança

A implementação do ÉGIDE requer um plano estruturado que traduza seus pilares em etapas concretas. Para facilitar o entendimento, a Tabela 3 apresenta a Matriz de Implementação, que organiza ações, ferramentas, responsáveis, prazos e custos.

Tabela 3. Matriz de Implementação

Etapa	Ação Principal	Ferramentas	Responsáveis	Prazo	Custos
1. Diagnóstico e Alinhamento	Mapear comportamento-alvo (inadimplência), barreiras cognitivas e oportunidades	Canvas de Intervenção; Behaviour Change Wheel (BCW)	Inteligência Fiscal (líder), TI, Comunicação	30 dias	Horas técnicas; eventual consultoria comportamental
2. Governança Ética e Algorítmica	Criar Comitê de Governança Ética; preencher Matriz Ética	Matriz de Governança Ética; registros LGPD	Jurídico (líder), TI, Compliance, Ouvidoria	30 dias	Reuniões internas; capacitação; pareceres jurídicos
3. Validação em pequena escala	Testes A/B em amostra de contribuintes; análise de resultados e efeitos colaterais	Protocolo de validação; planilhas; softwares estatísticos	Equipe de Dados & Análise (líder), Comunicação	60 dias	Desenvolvimento de mensagens; TI; licenças estatísticas

4. Implementação em larga escala	Expandir intervenção; registrar logs; supervisionar e auditar	Matriz RACI; sistemas de rastreabilidade; dashboards	TI & Arrecadação (líderes conjuntos); Comitê de Governança	90 dias	Infraestrutura tecnológica; auditoria externa
5. Transparência e Aprendizado Contínuo	Publicar relatórios; pesquisas de percepção; revisão anual da intervenção	Relatórios de impacto; indicadores de desempenho; pesquisas	Comunicação (líder), Ouvidoria, Jurídico, TI	Relatórios semestrais Revisão anual	Produção de relatórios; pesquisas; portal de transparência

Fonte: Elaboração própria.

Adicionalmente, o gestor pode fazer uso da Matriz RACI, para orientar o processo de responsabilização em cada pilar do ÉGIDE (Tabela 4). Essa matriz deve ser adaptada a realidade de cada entidade, para refletir os papéis e responsabilidades.

Tabela 4. Matriz RACI para a Implementação

Atividade	**Responsible (R)**	**Accountable (A)**	**Consulted (C)**	**Informed (I)**
Diagnóstico comportamental e alinhamento estratégico	Equipe de Inteligência Fiscal	Coordenação de Arrecadação	TI; Comunicação	Gabinete da SEFAZ
Elaboração e preenchimento da Matriz Ética	Jurídico	Comitê de Governança Ética	Compliance; Ouvidoria; TI	Gabinete; Controladoria
Planejamento e execução de testes A/B	Equipe de Dados & Análise	Coordenação de TI	Comunicação; Jurídico	Secretaria Executiva
Expansão da intervenção em larga escala	TI & Arrecadação	Coordenação Geral de TI	Auditoria Interna; Jurídico	Gabinete; Tribunal de Contas
Registro, rastreabilidade e auditorias periódicas	TI	Comitê de Governança	Auditoria Externa; Controladoria	Órgãos de controle
Publicação de relatórios e pesquisas de percepção	Comunicação	Coordenação de Comunicação	Ouvidoria; Jurídico; TI	Cidadãos; Sociedade civil

Fonte: Elaboração Própria. O termo RACI é um acrônimo conhecido nos modelos de TIC e reflete a indicação dos papeis de Responsável (Executor), Accountable (Reponsabilizado), Consultado (que pode opinar) e Informado (que não pode opinar).

6. Conclusões e contribuições

O presente artigo apresentou o ÉGIDE – Ética e Governança para Intervenções Digitais no Estado, um framework tecnológico concebido para orientar gestores públicos na aplicação responsável de *nudge*s digitais personalizados. A proposta surgiu do diagnóstico de fragilidades recorrentes em intervenções comportamentais, como opacidade algorítmica, fragmentação institucional, ausência de validação e escassez de mecanismos de transparência.

O ÉGIDE organiza essas lacunas em um conjunto de cinco pilares – diagnóstico, governança, validação, implementação e transparência – e os traduz em instrumentos práticos como o *Canvas de Intervenção Comportamental*, a *Matriz de Governança Ética*, protocolos de testes A/B e a *Matriz RACI*. Com isso, transforma princípios normativos e recomendações internacionais em um roteiro operacional adaptado ao contexto brasileiro.

A principal contribuição do framework está em sua replicabilidade. Embora o artigo tenha ilustrado sua aplicação no caso da inadimplência tributária, o ÉGIDE pode ser utilizado em diferentes áreas da administração pública, como saúde, educação e mobilidade urbana. Em todos esses domínios, oferece um modelo auditável e transparente, capaz de conciliar inovação tecnológica com conformidade regulatória e fortalecimento da confiança social.

Além dos ganhos potenciais de eficiência, a adoção do ÉGIDE promove benefícios institucionais de longo prazo: maior segurança jurídica, redução de riscos reputacionais e consolidação de práticas éticas no uso de dados e algoritmos. Assim, configura-se como um produto tecnológico aplicável, que contribui não apenas para o aumento da eficácia das políticas digitais, mas também para sua legitimidade democrática.

Referências

Behavioural Insights Team (2021). *Annual update report 2020–2021*. https://www.bi.team/publications/

Capasso, M., & Umbrello, S. (2022). Responsible nudging for social good: new healthcare skills for AI-driven digital personal assistants. *Medicine, Health Care and Philosophy*, *25*(1), 11-22. https://doi.org/10.1007/s11019-021-10062-z

DellaVigna, S., & Linos, E. (2020). RCTs to scale: Comprehensive evidence from two *nudge* units. *American Economic Review, 110*(12), 3830–3867. https://doi.org/10.3982/ECTA18709

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2012). MINDSPACE: Influencing behaviour through public policy. Institute for Government. https://www.instituteforgovernment.org.uk/publication/report/mindspace

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science, 6*, 42. https://doi.org/10.1186/1748-5908-6-42

Ogilvie, D., Craig, P., Griffin, S., Macintyre, S., & Wearmouth, L. (2015). A translational framework for public health research. *BMC Public Health, 9*(1), 116. https://doi.org/10.1186/1471-2458-9-116

OECD (2019). The innovation system of the public service of Brazil: An exploration of its past, present and future journey. *OECD Public Governance Reviews*. https://doi.org/10.1787/a1b203de-en

OECD/CAF (2022). The strategic and responsible use of artificial intelligence in the public sector of Latin America and the Caribbean. *OECD Publishing*. https://doi.org/10.1787/1f334543-en

Sugden, R. (2009). On Nudging: A Review of *Nudge*: Improving Decisions About Health, Wealth and Happiness by Richard H. Thaler and Cass R. Sunstein. *International Journal of the Economics of Business, 16*(3), 365–373. https://doi.org/10.1080/13571510903227064

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.

Van Ooijen, C., Ubaldi, B., & Welby, B. R. (2019). *A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance* (OECD Working Papers on Public Governance No. 33). OECD Publishing. https://doi.org/10.1787/09ab162c-en

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. Journal of Public Economics, 148, 14–31. https://doi.org/10.1016/j.jpubeco.2017.02.003

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505–523. https://ssrn.com/abstract=2972505

Zhao, J., Arya, V., Gupta, M., & Kose, U. (2021). Data-driven nudging for social good: Applications and challenges. *Information Systems Frontiers, 23*(5), 1171–1188. https://doi.org/10.1007/s10796-020-10008-2