Estratégia de Implantação da Inteligência Artificial pelas Defensorias Públicas do Brasil

Contribuição Data Privacy Brasil

Elaborado por

Johanna Monagreda
Eduardo Mendonça
Pedro Saliba
Mariana Rielli

DataPrivacyBR Research

Link da consulta:
Participa + Brasil - Estratégia de Implantação da Inteligência Artificial pelas Defensorias Públicas do Brasil

No dia 21 de dezembro de 2023, as Defensorias Públicas dos Estados e do Distrito Federal, representadas pelo Conselho Nacional das Defensoras e Defensores Públicos-Gerais (CONDEGE), e a Defensoria Pública da União iniciaram um processo de consulta pública em interesse de coletar subsídios de diferentes setores da sociedade para compor a Estratégia de implantação da inteligência artificial pelas Defensorias Públicas do Brasil.

A Data Privacy Brasil elaborou um contributo que reflete o compromisso da instituição com os direitos fundamentais, e os aprendizados de mais de 4 anos de colaboração e parceria com as Defensorias Públicas através dos projetos Expandindo o papel das Defensorias Públicas na proteção de dados no Brasil e Construindo diálogos e formando lideranças populares em direitos digitais junto às Defensorias Públicas Estaduais

A tecnologia de IA pode vir a ser uma ferramenta útil para facilitar algumas tarefas operativas das Defensorias Públicas, e assim tornar mais eficiente o papel da instituição como órgão efetivador do acesso à justiça para a população mais vulnerável. Contudo, múltiplos casos concretos e diversas pesquisas alertam sobre os impactos sociais, humanos e ambientais atuais, previsíveis ou imagináveis do uso da tecnologia de inteligência artificial.

Discutir os impactos com a cidadania é indispensável para garantir a transparência e legitimidade do processo. Daí a relevância desta iniciativa das DPs em promover uma consulta pública para definir de forma participativa parâmetros éticos e técnicos a partir do qual possa ser desenhado o sistema de governança para a implementação de inteligência artificial (IA).

A continuação segue nossa contribuição a esse debate, na expectativa de que a estratégia de IA a ser seguida pelo órgão de justiça adote um enfoque de ***regulamentação e gestão orientada por direitos, e embasada prioritariamente na proteção e defesa da população mais desfavorecida***, de modo a garantir que o emprego de IA nas atividades administrativas não prejudique a função principal de assistência jurídica e promoção dos direitos humanos. Nossas recomendações, portanto, giram em torno a adoção dos princípios de governança participativa e democrática, a implementação de mecanismos de avaliação de riscos e impactos, e supervisão humana dos processos automatizados.

Quais riscos, desafios e impactos sistêmicos devem ser considerados na implementação de sistemas de inteligência artificial pelas Defensorias Públicas? Que estratégias podem ser adotadas para eliminação ou mitigação desses riscos?

As Defensorias Públicas precisam implementar estratégias para que o emprego de IA em suas atividades administrativas não prejudique sua função principal de assistência jurídica e promoção dos direitos humanos. Essas estratégias incluem governança participativa e democrática, a implementação de mecanismos de avaliação de riscos e impactos e supervisão humana do processo automatizado.

A incorporação de uma metodologia de avaliação de riscos e impactos, que pode ser inspirada em modelos atualmente em discussão, como a Avaliação de Impacto Algorítmico (AIA), permite identificar e mitigar riscos, que são potenciais prejuízos conhecidos, previsíveis e especulativos às liberdades fundamentais e aos direitos individuais e coletivos. Para um órgão público com mandato constitucional de promoção e defesa de direitos humanos, a AIA permite a identificação prévia de riscos inaceitáveis ou altos, a decisão pela proibição em defesa dos direitos humanos e/ou liberdades fundamentais, e a identificação de tecnologias com riscos menores que, com medidas de prevenção, mitigação e monitoramento, poderiam ser implementadas.

É importante garantir, ainda, o direito de revisão dos processos e decisões automatizadas, estabelecer mecanismos de participação cidadã, controle social, prestação de contas e auditoria.

A implementação da IA, se não acompanhada de tais medidas, desconsidera que os direitos humanos são universais, inalienáveis, indivisíveis, interdependentes e interrelacionados e negligencia o interesse público e o acesso efetivo à justiça pelos mais vulneráveis, indo de encontro com a própria missão e atribuição legal do órgão. A tecnologia não pode se tornar uma barreira ao acesso à justiça, portanto é importante avaliar constantemente que a tecnologia não funcione como mecanismo de exclusão a pessoas hipossuficientes e que as pessoas consigam, de fato, usar a tecnologia.

Quais aspectos relacionados à proteção de dados e transparência devem ser considerados pelas Defensorias Públicas na formulação e implementação de sua estratégia de inteligência artificial? Como esses aspectos podem ser articulados e mobilizados para assegurar a proteção de dados e transparência, inclusive a transparência algorítmica?

O acelerado e massivo fluxo e processamento de dados com a implementação de IA, a opacidade algorítmica associada à tecnologia, o grau de autonomia envolvida na IA de *machine learning* e *deep learning* colocam desafios à proteção de dados pessoais na implementação de IA.

As Defensorias Públicas, em seu duplo papel como agente controlador de dados pessoais de pessoas especialmente vulneráveis, e como órgão de justiça responsável pela proteção de dados da cidadania frente a terceiros, deverá garantir a proteção dos dados pessoais dos titulares e adequar o processo de implantação de IA às exigências da LGPD para a implementação de IA. Torna-se indispensável a supervisão em todo o processo de implementação da IA pela figura de Encarregado de Proteção de Dados, bem como pela Autoridade Nacional de Proteção de Dados, como garantia aos direitos dos titulares dos dados pessoais.

É recomendável a elaboração de instrumentos de governança de dados como a política de privacidade e o relatório de impacto à proteção de dados pessoais; o estabelecimento de sistemas de autenticação, mecanismos de segurança dos dados, de prestação de contas e auditoria, com especial atenção para o tratamento de dados sensíveis, e informar sobre o compartilhamento de dados pessoais entre órgãos da administração pública, as garantias de segurança da informação frente a riscos de vazamento, ou desvio de finalidade, especialmente frente aos perigos de vigilância excessiva que envolvem o uso de algoritmos de *machine-learning* e *deep-learning* que podem responder a diferentes finalidades.

Os princípios da proteção de dados, bem como transparência, responsabilização e prestação de contas podem tornar o processo de implementação de IA mais justo. É recomendável a incorporação de elementos de explicabilidade da IA e do processo de implementação da tecnologia no que se refere aos dados envolvidos, especialmente relevantes quando se trata de IA generativa ou programada mediante aprendizado de máquina (*machine learning* ou *deep-learning*).

Como sistemas de inteligência artificial podem promover ou agravar violações de direitos humanos, considerando, inclusive, o racismo algorítmico e outros vieses potenciais da tecnologia? Quais medidas podem ser adotadas nas diferentes etapas dos projetos para assegurar que as soluções de inteligência artificial não reproduzam preconceitos e, ao contrário, possam contribuir para a promoção dos direitos humanos a partir da atuação da Defensoria Pública?

Sistemas de IA têm o potencial de agravar violações de direitos humanos, uma vez que a capacidade de processar *Big Data* aumenta os riscos de vigilância excessiva, perda da privacidade e apropriação indevida de dados sensíveis. Eles podem datificar a vida cotidiana de grupos historicamente vulnerabilizados, automatizar vieses e injustiças associadas ao racismo, sexismo e desigualdades socioeconômicas, facilitando o perfilamento racial e discriminação e produzem impactos negativos no livre desenvolvimento da pessoa humana e no princípio da dignidade humana, entre outros.

Além do alcance e a própria finalidade da tecnologia, a utilização de bases de dados enviesadas, a utilização de proxys raciais ou de gênero com fim de discriminaçao ilícita; a programação, intencional ou não, do algoritmo; ou o próprio aprendizado da máquina na interação humana, no caso de tecnologia *machine learning*, tem sido apontados como possíveis causas de resultado discriminatório e contrário aos direitos humanos da IA.

Para proteger os direitos, é crucial seguir exemplos globais e locais de regulação e uso desses sistemas, como medidas protetivas definidas por organizações internacionais como a ONU e UNICEF, além de organismos como a Access Now. A inclusão dos mais vulneráveis na aplicação de sistemas, juntamente com estratégias de governança para o desenvolvimento seguro dessas tecnologias, são fundamentais.

Localmente, exemplos incluem o PL 2.338/2023, que reconhece grupos especialmente vulneráveis e garante o direito à correção de vieses algorítmicos e a revisão humana. Medidas de mitigação, como abordagem baseada em princípios de não-discriminação, responsabilização de atores em casos de discriminação algorítmica, uso de software aberto e Avaliação de Impacto Algoritmico também são essenciais, assim como a inclusão de atores sociais no desenvolvimento de políticas públicas.

Que elementos devem ser considerados para a promoção da educação e qualificação de defensores e demais servidores no tema da inteligência artificial, pelos centros de estudo, fundações e escolas das Defensorias Públicas e outros caminhos complementares para formação e capacitação continuada?

Existem três camadas a serem consideradas: técnica, de compreensão sobre como as ferramentas funcionam; social, com as implicações contextuais sobre seu uso; e procedimental, relacionada à administração pública.

As Defensorias precisam capacitar seus quadros para entender os conceitos da IA, incluindo algoritmos, aprendizado de máquina, redes neurais, entre outros. É recomendado fornecer cursos introdutórios sobre IA e tecnologias relacionadas, considerando seus aspectos técnicos e uma compreensão geral sobre como são obtidos os resultados das ferramentas. Esse tipo de estudo pode ser concomitante a treinamentos de usos das plataformas.

Na camada social, deve-se considerar imposições legais e éticas da aplicação da IA em áreas como justiça criminal, saúde, educação e assistência social, compreendendo como ferramentas de IA aplicadas pelo setor público e privado podem violar direitos individuais e coletivos. As perspectivas das DPEs devem ser consideradas tanto na atividade fim quanto em eventuais usos desses sistemas para gestão de processos judiciais e demandas administrativas. O trabalho interdisciplinar garante uma série de perspectivas a partir de pesquisas e interlocução com atendidos e atendidas pela instituição. Destaca-se o trabalho da assistência e ciências sociais nesse campo, uma vez que a análise jurídica, por si só, não basta para compreensão do fenômeno. Essas reflexões devem ser validadas por profissionais técnicas para destacar os limites e possibilidades da ferramenta.

A camada procedimental deve levar em conta não apenas o treinamento para uso lícito e ético sobre IA, mas também considerar o desenvolvimento dessas ferramentas. Havendo contratações externas, é importante destacar no procedimento administrativo – licitatório ou outras formas contratuais – requisitos técnicos e documentação dos processos para eventual auditoria. Importante realizar a documentação das práticas com Relatórios de Impacto com metodologias validadas por organismos legais e científicos.

Quais recomendações podem ser feitas para a melhor governança da estratégia unificada das Defensorias Públicas para adoção de soluções de inteligência artificial?

Os mecanismos de governança da IA devem incorporar uma gestão ética e orientada pela promoção e defesa dos direitos humanos em todas suas etapas, transparência, explicabilidade algorítmica, participação e controle cidadão, com mecanismos adequados de avaliação de impactos e prestação de contas pública.

Parte da estratégia de governança implica a identificação das etapas e/ou tarefas centrais para a efetivação do acesso à justiça e que - de acordo com a natureza do órgão, a competência constitucional e as características da população atendida – se automatizadas colocariam em risco a própria missão da instituição; a partir disso, estabelecer parâmetros éticos compartilhados para a implementação de sistemas de IA e para a capacitação e atualização do recurso humano; atualizar a Infraestrutura tecnológica; estabelecer mecanismos de segurança da informação; incluir mecanismos de fácil identificação da autenticidade e veracidade do conteúdo digital produzido pela instituição; explicitar o ciclo de vida da IA, os agentes envolvidos (desenvolvedores, importadores, operadores), as obrigações, as responsabilidades e a análise prévia em casos em que a implementação da tecnologia possa resultar em vulneração de direitos ou princípios, as medidas de mitigação e os meios de reparação do dano, caso haja.

A estratégia de governança precisa ser capaz de lidar com a complexidade e dinamismo das tecnologias de IA, especialmente de IA generativa, onde a capacidade de deep-learning pode distanciar o funcionamento da tecnologia da finalidade inicial programada, e inclusive produzir riscos difíceis de prever. O grau de autonomia da tecnologia não exime da responsabilização humana.

É recomendável estabelecer mecanismos de avaliação contínua do próprio modelo de governança (procedimentos organizacionais, estruturas de responsabilização dos agentes internos e externos), do modelo algorítmico (programação, desempenho, taxa de erro), do impacto nos direitos humanos, na cidadania e no acesso à justiça.

Há outras sugestões ou comentários a serem feitos, relacionados a aspectos não explorados no texto e nas perguntas, que possam contribuir para a formulação e implementação da estratégia de inteligência artificial das Defensorias Públicas?

ALGUNS DOCUMENTOS SUGERIDOS

AccessNow. “Radiografía normativa: ¿dónde, qué y cómo se está regulando la inteligencia artificial en América Latina?: Informe de políticas públicas de IA en América Latina”, 2024.

Bioni; Mesquita; Monagreda; Zanatta (Orgs). Construindo caminhos para a justiça de dados no Brasil: o papel das Defensorias Públicas na proteção de dados pessoais, Data Privacy Brasil, 2022.

Bioni, B, Garrote, e Guedes. “Temas centrais na Regulação de IA: O local, o regional e o global na busca da interoperabilidade regulatória”. Data Privacy Brasil, 2023.

Gonçalves, A; Torres, e Melo (Orgs). Inteligência artificial e algoritmos - Desafios e oportunidades para os media, LabCom, 2024.

Lima, T. Racismo algorítmico. Coleção Panorama. Rio de Janeiro, RJ: CESec, 2023.

Mantelero, A. Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI. Information Technology and Law Series, 36. Asser Press, 2022.

Monagreda, J K. “Por que falar de raça quando falamos de dados pessoais, inteligência artificial e algoritmos?” Em Inteligência artificial e algoritmos - Desafios e oportunidades para os media, organizado por Adriana Gonçalves, Luisa Torre, e Paulo Victor Melo, 2024.

Noble, Safiya. Algorithms of oppression: how search engines reinforce racism. New York: New York University Press, 2018.

O’Neil, C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Broadway Books, 2017.

Pacheco, R. Projeto de Lei. Dispõe sobre o uso da Inteligência Artificial, Pub. L. No. 2338 (2023). https://www25.senado.leg.br/web/atividade/materias/-/materia/157233.

Silva, T. Racismo algorítmico: inteligência artificial e discriminação nas redes digitais. Edições Sesc SP, 2022.

UNESCO. “Recomendação sobre a Ética da Inteligência Artificial”, 2022.

Declaración de Montevideo sobre IA

Estratégia Brasileira para Transformação Digital

Estratégia Brasileira de Inteligência Artificial -EBIA-

Encuesta - Herramienta de autoevaluación ética para el sector público

ETHICALLY ALIGNED DESIGN