

RECOMENDAÇÕES
DE GOVERNANÇA

USO DE
INTELIGÊNCIA
ARTIFICIAL
PELO PODER
PÚBLICO

REALIZAÇÃO



DIRETOR-EXECUTIVO

Manoel Galdino

DIRETORA DE OPERAÇÕES

Juliana Sakai

COORDENADORA DO PROJETO

Tamara Burg

PESQUISA

Jonas Coelho

Tamara Burg

TEXTO

Juliana Sakai

Manoel Galdino

Tamara Burg

DIAGRAMAÇÃO

Marina Atoji

www.transparencia.org.br

FINANCIAMENTO



PARCERIA

Northwestern University

COLABORAÇÃO

Artigo 19

Conectas Direitos Humanos

Instituto Brasileiro de Defesa do

Consumidor (IDEC)

Instituto de Defesa do Direito de Defesa

(IDDD)

Instituto de Estudos da Religião (ISER)

Instituto de Referência em Internet e

Sociedade (IRIS)

Instituto de Tecnologia e Sociedade (ITS)

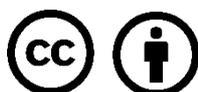
Instituto Igarapé

Instituto Socioambiental (ISA)

Minas Programam

Mulheres Negras Decidem

PretaLab



CC-BY

Este trabalho está sob a licença [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/). Mediante atribuição de crédito à organização autora, pode ser copiado e redistribuído em qualquer suporte ou formato; remixado e adaptado para qualquer fim, inclusive comercial (nestes casos, as alterações feitas devem ser indicadas). Fevereiro/2020.

SUMÁRIO

INTRODUÇÃO	2
ANÁLISE DOS CASOS DE USO	4
ANÁLISE DOS RISCOS A DIREITOS.....	8
SEGURANÇA PÚBLICA.....	14
1. Reconhecimento facial.....	14
2. Processamento de linguagem natural.....	16
RECOMENDAÇÕES DE GOVERNANÇA.....	17
1. Bases de dados representativas e apropriadas para o contexto.....	17
2. Necessidade de supervisão humana como salvaguarda para a revisão de decisões automatizadas.....	19
3. Efetiva proteção dos dados pessoais do cidadão.....	21
4. Transparência e explicabilidade dos sistemas	23
ANEXO 01	27
Catálogo de ferramentas de IA usadas pelo setor público brasileiro	27

INTRODUÇÃO

Este documento apresenta as recomendações da **Transparência Brasil** para o desenvolvimento e uso de tecnologias relacionadas à Inteligência Artificial (IA)¹ pelo setor público brasileiro, elaboradas de forma colaborativa com as seguintes² organizações da sociedade civil, atuantes na promoção de diferentes causas e direitos: **Artigo 19**; **Conectas Direitos Humanos**; **Instituto Brasileiro de Defesa do Consumidor (IDEC)**; **Instituto de Defesa do Direito de Defesa (IDDD)**; **Instituto de Estudos da Religião (ISER)**; **Instituto de Referência em Internet e Sociedade (IRIS)**; **Instituto de Tecnologia e Sociedade (ITS)**; **Instituto Igarapé**; **Instituto Socioambiental (ISA)**; **Minas Programam**; **Mulheres Negras Decidem** e **PretaLab**³.

A implementação responsável das tecnologias que usam técnicas de IA para melhorar a prestação de serviços públicos pode trazer benefícios para a sociedade. No entanto, é necessário considerar que os usos dessas tecnologias também podem trazer impactos negativos ao exercício de direitos pela sociedade, como o direito à privacidade, à proteção contra a

¹ A OCDE define sistemas de inteligência artificial como “a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.” OECD Legal Instruments, Recommendation of the Council on Artificial Intelligence, 2019. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

² Organizações listadas em ordem alfabética.

³ Agradecemos também o apoio das seguintes organizações e pessoas: Controladoria-Geral da União (CGU), Ministério da Ciência, Tecnologia e Inovação (MCTI), Centro de Estudos sobre Tecnologias Web (Ceweb.br), CodingRights, Data Privacy Brasil, InternetLab, Bruno Kunzler e Daniel Trielli.

discriminação, ao acesso à justiça, e às liberdades de expressão, associação e reunião, dentre outros, conforme será apontado neste documento.

Ademais, a falta de parâmetros legais deixa em aberto uma lacuna jurídica, regulatória e ética, com as más consequências que o uso de sistemas de IA sem governança pode trazer.

O trabalho aqui desenvolvido oferece um panorama de ferramentas de IA usadas pelos Poderes Executivo, Legislativo e Judiciário brasileiros para atingir diferentes finalidades dentro da esfera governamental, bem como uma análise dos possíveis impactos negativos a direitos e uma representação das principais preocupações apontadas pela sociedade civil sobre a utilização de tais tecnologias.

Este documento busca sintetizar a análise de impactos negativos a direitos fundamentais realizada por representantes das organizações da sociedade civil acima mencionadas, para, ao final, oferecer recomendações de governança quanto à aplicação de sistemas de IA no setor público.

ANÁLISE DOS CASOS DE USO

Inicialmente, foi feito um mapeamento de algoritmos de IA e seus casos de uso no poder Executivo brasileiro. A metodologia para obter essas informações consistiu em um questionário enviado a órgãos da administração pública direta e indireta do Executivo federal, pedidos de acesso à informação e busca em transparência ativa. Para tanto, um questionário foi enviado. O resultado é um catálogo de ferramentas de IA usadas pelo setor público brasileiro (ANEXO 01).

O questionário, solicitando informações sobre a utilização de tecnologias de IA em suas iniciativas, foi elaborado em parceria com a Controladoria-Geral da União (CGU), o Ministério da Ciência, Tecnologia e Inovação (MCTI) e o Centro de Estudos sobre Tecnologias Web (Ceweb.br) do Núcleo de Informação e Coordenação do Ponto BR (NIC.br). Em setembro de 2020, enviamos o questionário aos 319 órgãos do Executivo Federal que fazem parte da lista dos órgãos federais do Sistema Eletrônico de Informações ao Cidadão (e-SIC).

Adicionalmente, [pedidos de acesso à informação](#) foram feitos aos órgãos do Legislativo e Judiciário, requerendo as mesmas informações do questionário⁴. Neste sentido, foram realizados pedidos de acesso à informação aos seguintes órgãos: Senado Federal, Câmara dos Deputados, Tribunal de Contas da União (TCU), Conselho Nacional de Justiça (CNJ), Superior Tribunal de

⁴Todos os pedidos, bem como as respostas, podem ser acessados no projeto Achados e Pedidos, que contém pedidos de LAI, em:

<http://www.achadosepedidos.org.br/usuarios/tburg>

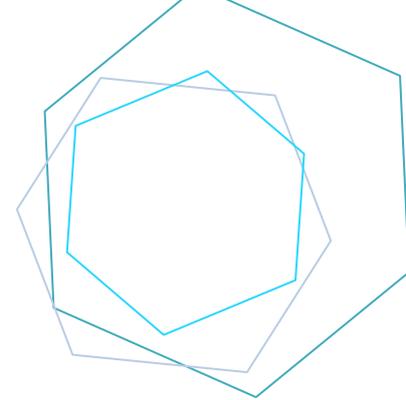
Justiça (STJ), Supremo Tribunal Federal (STF) e Tribunal Superior do Trabalho (TST).

Finalmente, desenvolvemos, internamente, com o apoio da Northwestern University, um algoritmo que busca encontrar, de maneira automatizada, ferramentas de IA usadas pelo setor público. O algoritmo foi desenvolvido por meio dos resultados do Google nas buscas por palavras-chaves de termos associados à inteligência artificial, aprendizado por máquina e modelos preditivos. A busca foi filtrada para exibir apenas resultados presentes em sites governamentais brasileiros terminados em ".gov.br", ".leg.br", ".jus.br" e ".mp.br".

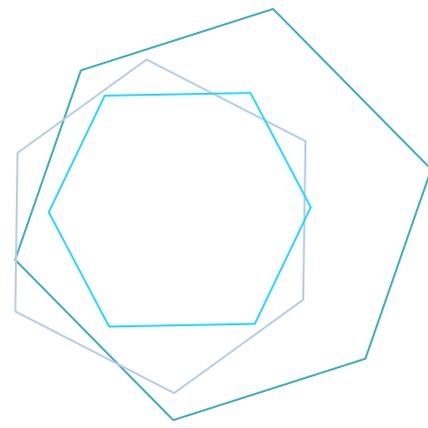
Com base nesses resultados, foram manualmente avaliados 6.195 endereços de sites governamentais (URLs) de acordo com o conteúdo que apresentavam para julgar se estavam relacionados ao tema ou não. Em seguida, todas as páginas tiveram seu conteúdo em texto extraído e, usando o pacote Scikit-Learn em Python, foi desenvolvido um modelo de regressão logística visando prever termos associados com resultados positivos. O código utilizado pode ser encontrado em nosso repositório no Github⁵. O modelo visa contribuir para que a sociedade civil possa monitorar novos casos de uso de tecnologias de IA que venham a ter menção em sites governamentais.

Para realizar uma análise multissetorial que cobrisse diversos campos, organizações da sociedade civil de diferentes áreas de atuação contribuiriam no processo de avaliação de riscos a direitos.

⁵ Link para o repositório público: <https://github.com/Transparencia-Brasil/algoritmos-brasil>. O algoritmo desenvolvido ainda será testado para garantir que não ocorram vieses em seu funcionamento e todos os testes estarão disponíveis no repositório.



Até o momento, foram mapeadas 44 ferramentas de IA usadas por órgãos governamentais, agrupadas aqui em duas dimensões: a primeira dimensão separa as ferramentas segundo sua utilização para tomada de decisão ou não e a segunda considera se o público-alvo é interno ou externo ao poder público (usuários finais).



Na primeira dimensão, classificamos uma ferramenta como usada para tomada de decisão ou apoio na tomada de decisão se ela toma uma decisão autonomamente ou se foi criada com o propósito de apoiar uma decisão humana. Por exemplo, a ferramenta Bem-te-vi, usada pelo Tribunal Superior do Trabalho para classificação de processos e previsões sobre a tramitação do processo nos gabinetes, pode afetar o exercício do direito fundamental de acesso à Justiça, ao devido processo legal e ao dever legal de motivação das decisões judiciais caso automatize decisões que deveriam ser fruto de uma avaliação humana.

E classificamos o algoritmo como não envolvendo decisão se a ferramenta é utilizada apenas para resolver problemas internos de gestão que não envolvam tomada de decisão, como a automação de processos rotineiros. A título de exemplo, a ferramenta Victor, usada pelo Supremo Tribunal Federal, que objetiva simplificar o reconhecimento de padrões em textos jurídicos apresentados perante o STF, ao analisar os recursos extraordinários e identificar quais estão vinculados a determinados temas de repercussão geral, não foi criada para auxiliar alguma decisão judicial diretamente, mas apenas indiretamente, na medida em que torna os processos mais ágeis.

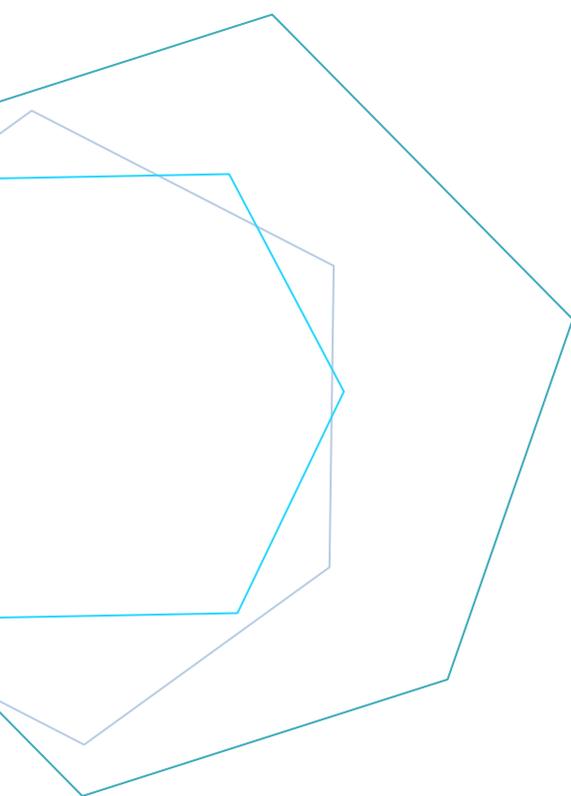
A segunda dimensão classifica os algoritmos segundo o público-alvo: interno (isto é, os próprios agentes governamentais que interagem com a ferramenta) ou externo

(cidadãos, empresas e outras entidades que são impactadas pelo uso da ferramenta).

QUADRO 1: CLASSIFICAÇÃO DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL (IA) SEGUNDO USO PARA TOMADA DE DECISÃO E PÚBLICO-ALVO.

	COM TOMADA DE DECISÃO	SEM TOMADA DE DECISÃO	TOTAL
Público interno	20 ferramentas	16 ferramentas	36
Externo	8 ferramentas	0 ferramentas	8
Total	28	16	44

Com relação a envolver tomada de decisão ou não, 28 ferramentas (64%) são de apoio a processos de tomada de decisão governamental, enquanto 16 ferramentas (36%) são utilizadas para auxiliar em demandas internas sem processo decisório envolvido. Com relação ao público-alvo, 36 ferramentas (82%) são para uso interno dos servidores públicos que atuam no órgão governamental e 8 (18%) interagem diretamente com usuários externos (cidadãos/público em geral).



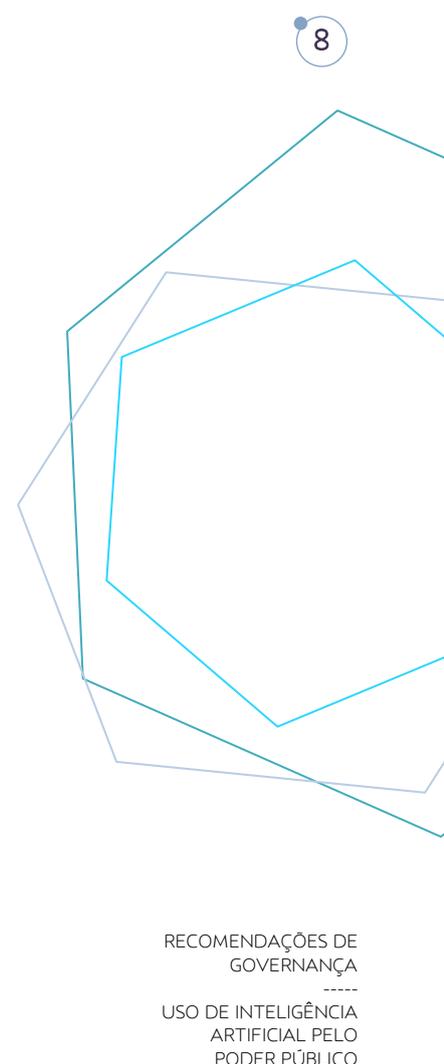
ANÁLISE DOS RISCOS A DIREITOS

Ao se discutir e propor recomendações de governança para o uso de algoritmos de IA, é importante considerar a avaliação de riscos envolvendo ameaças reais e potenciais a direitos e ao espaço cívico, buscando alinhar promoção de inovação e tecnologia com responsabilidade pública e transparência.

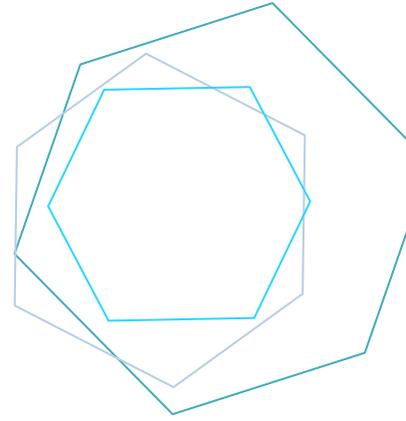
Nesta perspectiva, sugerimos classificar o potencial impacto a direitos que determinada ferramenta de IA pode causar em casos concretos com base no “output” ou em seus resultados, isto é, no resultado que ela foi desenhada para entregar ou, ainda, caso ocorra um possível erro em seu funcionamento. Esta análise completa encontra-se na **estrutura de avaliação de riscos a direitos e de transparência no uso de algoritmos de inteligência artificial no setor público**⁶, proposta pela Transparência Brasil, com participação de especialistas no tema e organizações da sociedade civil.

Tal classificação é necessária, pois algoritmos de IA podem ser utilizados para atingir diferentes finalidades dentro da esfera governamental e, da mesma forma, uma vez que a natureza e complexidade das ferramentas utilizadas são diversas entre si, elas podem ocasionar riscos a direitos de diferentes aspectos e gravidades.

⁶ A estrutura de avaliação de riscos será publicada em breve no site da Transparência Brasil.



Uma correta identificação inicial do risco que o algoritmo representa a diferentes direitos permite mitigar esses riscos. O desenho e a implementação do algoritmo podem ser pensados de antemão para prevenir esses riscos e torna-se mais fácil haver um acompanhamento do seu funcionamento e resultados. Além disso, pode-se avaliar a pertinência de se ter um aparato regulatório específico para todos os diferentes sistemas tecnológicos de IA.



A partir do cruzamento das tipologias, conforme pode ser observado no quadro 1, temos três tipos de ferramentas atualmente em uso pelo poder público brasileiro de acordo com o mapeamento da Transparência Brasil: i) 20 ferramentas para apoio a tomada de decisão direcionadas para os próprios órgãos públicos; ii) 8 ferramentas de decisão direcionadas para o público externo e; iii) 16 ferramentas para aperfeiçoar processos internos dos órgãos, sem envolver tomada de decisão.

As ferramentas de apoio a processo de tomada de decisão utilizadas internamente pelos órgãos governamentais constituem o maior número de casos mapeados - e são as que mais geram preocupação para a sociedade civil - pois estas ferramentas ajudam servidores de órgãos governamentais a tomarem determinadas ações que impactam a vida das pessoas e, de forma direta ou indireta, o exercício de direitos fundamentais.

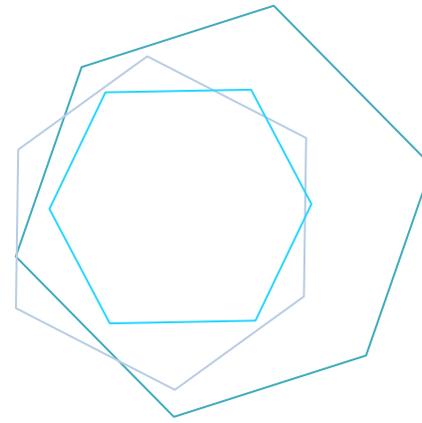
O principal risco apontado diz respeito ao banco de dados de treinamento e critérios a serem utilizados pelos modelos preditivos e de classificação automatizados, usados pelos órgãos governamentais, que podem levar à reprodução de uma discriminação social já pré-existente, afetando, principalmente, parte da população socialmente mais vulnerável.

Um exemplo é o uso da ferramenta PalasNET pelo Departamento de Polícia Federal. De acordo com o órgão, o sistema é utilizado para registro de informações da fase sigilosa das investigações e utiliza técnicas de reconhecimento de imagens, reconhecimento facial e perfil criminal.

Ainda de acordo com o órgão, um banco de dados criminais foi usado como variável para treinar a ferramenta. Se o banco de dados não for representativo da população, pode gerar vieses discriminatórios com maior chance de falso positivos para determinados grupos, como negros e pessoas de baixa renda ou das periferias urbanas. Como o sistema criminal e judicial brasileiro é altamente discriminatório, essa possibilidade se torna na verdade uma probabilidade bastante alta e preocupante.

Outra ferramenta usada em contexto diferente mas com preocupações de natureza semelhante é a ferramenta Weka, da Universidade Federal de Santa Maria, empregada para análise de predição de abandono no ensino superior. De acordo com o órgão, dados acadêmicos são usados como variáveis para treinar a ferramenta.

O uso desta ferramenta pode impactar negativamente o direito à educação caso seja utilizada para tomada de ações que penalizam aqueles classificados com alta probabilidade de abandono (como negação de benefício de auxílios estudantis). O impacto negativo pode ser criado ao incorporar fatores discriminatórios para a análise, como, por exemplo, considerar dificuldades financeiras ou a maternidade como fatores que aumentem a probabilidade de abandono. Isso pode se tornar discriminatório, prejudicando mulheres, negros e a população de baixa renda no acesso à educação. Ademais, pode criar um círculo vicioso, na medida em que a negação a esses benefícios

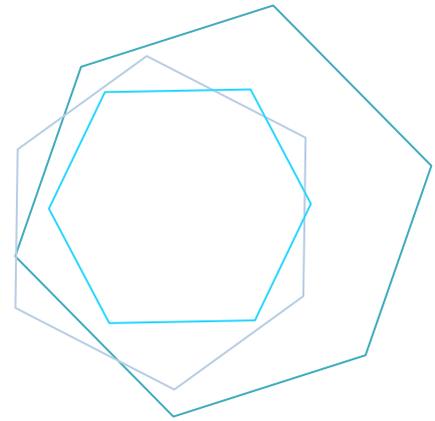


pode contribuir para o próprio abandono escolar, “confirmando” as previsões da ferramenta.

O segundo tipo de ferramenta mapeada corresponde às ferramentas de apoio à tomada de decisão e que interagem diretamente com usuários externos (cidadãos/público em geral). Este é o caso dos *chatbots*, muitas vezes desenvolvidos para ajudar a direcionar o usuário de um serviço público com vistas a reduzir recursos humanos e tornar o serviço público mais eficiente.

Os riscos apontados com o uso destas ferramentas, de modo geral, dizem respeito ao acesso ao serviço por pessoas cuja inclusão digital vem sendo realizada de forma precária, incompleta ou inadequada, como analfabetas, imigrantes, deficientes⁷ ou sem acesso estável à internet. O *chatbot* é a porta de entrada para a obtenção de serviços e para o exercício de uma série de direitos. Assim, mesmo em seu bom uso o sistema pode impactar negativamente cidadãos por não considerar usos e costumes locais, impedindo o acesso a serviços públicos.

Por exemplo, o *chatbot* para realizar a triagem de pacientes com sintomas de Covid-19 usado pelo Hospital Universitário da Universidade Federal do Maranhão pode ocasionar impactos negativos tanto por meio de vieses e problemas de triagem gerados (ex.: não reconhecimento de uma combinação específica de sintomas como um indicativo de infecção por coronavírus) além de impactar negativamente minorias sociais



⁷ É importante notar que o risco não deriva da condição ou característica da pessoa, como o caso de pessoas deficientes deixa talvez mais evidente, já que elas podem ser excluídas do acesso aos serviços se essas ferramentas não contemplarem preocupações com acessibilidade desde o seu desenvolvimento.

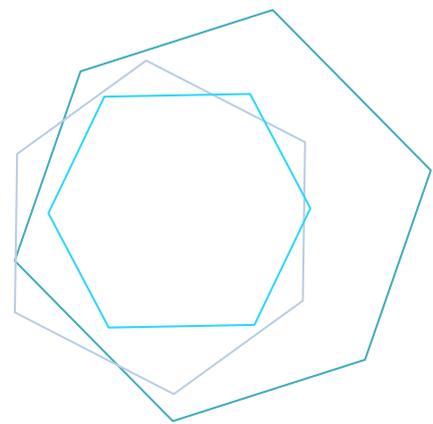
que não possuem um bom domínio da linguagem escrita, trazendo riscos à saúde do cidadão.

A depender dos critérios utilizados, as triagens e classificações de risco podem ser discriminatórias, favorecendo pessoas brancas e de classes sociais mais privilegiadas.

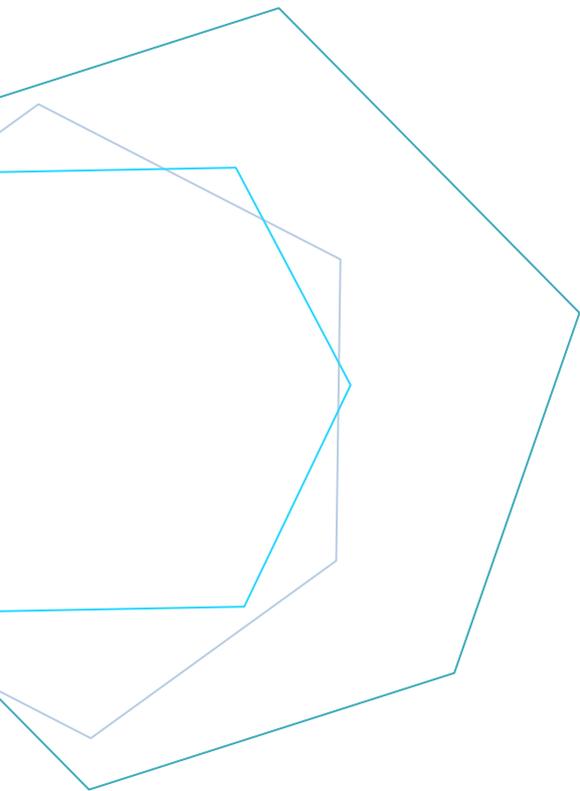
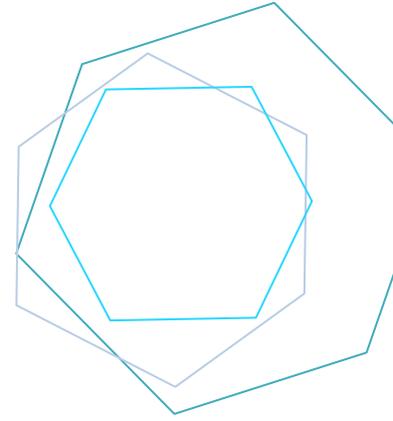
Ademais, o próprio direito à saúde pode ser afetado, já que a ferramenta funciona a partir de uma sequência pré-definida de comandos automatizados que, com base em dados pessoais e não pessoais, chega a conclusões que podem impactar significativamente a vida do cidadão, que não consegue pedir explicações imediatas sobre como tais conclusões foram tomadas.

Por fim, o terceiro caso mapeado é o uso de ferramentas de IA voltadas a auxiliar na simplificação de processos de análise do setor público. Ou seja, seu uso busca resolver um problema do próprio servidor público, ou da capacidade da administração pública de processar todas as demandas. Por essa razão, também possui um risco mais baixo de violação de direitos e, se bem utilizada, a IA pode promover melhorias nos processos administrativos, aumentar a eficiência e diminuir gastos públicos.

Como exemplos, podemos citar a ferramenta, ainda em desenvolvimento, a ser utilizada pela Agência Nacional de Vigilância Sanitária, para facilitar a análise das contribuições a consultas públicas publicadas por meio de agrupamento de informações de texto ou, ainda, a ferramenta para classificação de demandas dos cidadãos recebidas pela ouvidoria do Tribunal de Contas da União para fins de comparação e facilitação de atendimento.



Embora essas ferramentas pareçam ter menor potencial de impacto discriminatório, o seu uso pelo setor público deve se valer da exigência de transparência e das demais regras aplicáveis à administração pública, como, por exemplo, os princípios da moralidade, legalidade, impessoalidade, publicidade e eficiência, elencados no artigo 37 da Constituição Federal.



SEGURANÇA PÚBLICA

Ferramentas de segurança pública ocupam lugar proeminente em muitas discussões sobre regulação de uso de IA em muitos lugares do mundo⁸. Embora este documento não considere que alguns direitos sejam mais importantes que outros, dada a centralidade do uso de IA em segurança pública nos tempos atuais, especialmente em uma sociedade racista como a brasileira, faz sentido abordar especificamente este ponto neste documento. Assim, apresentamos dois exemplos concretos de tecnologias de Inteligência Artificial usadas pelo governo para fins de segurança pública, seus riscos e impactos que devem ser potencialmente levados em consideração em seu uso e implementação.

1. Reconhecimento facial

O uso generalizado de sistemas de reconhecimento facial para fins de segurança pública. -- por exemplo em locais públicos -- incide sobre toda a população que transita pelo local, potencialmente ferindo a presunção de inocência de todos. Ademais, quando não analisados para detecção de possíveis vieses, podem impactar negativamente o princípio da não-discriminação. Um exemplo mapeado é a ferramenta AI Desk usada pelo Departamento de Polícia Federal. A Desk utiliza as técnicas de reconhecimento facial e classificação de imagens

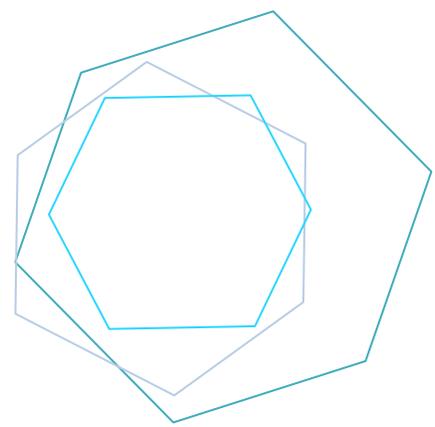
⁸ A Anistia Internacional, por exemplo, lançou uma campanha para banir seu uso: <https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/>

para i) auxiliar na identificação de rostos, ii) classificar rostos em grupos de idades, iii) identificar objetos em cenas e iv) estimar a probabilidade de determinada imagem conter nudez.

Nossa análise indicou que o uso de reconhecimento facial para fins de vigilância sempre impacta negativamente o princípio da presunção de inocência, mesmo que, em um cenário ideal, funcionasse sem vieses. Isso porque o uso da tecnologia inverte a lógica do direito penal em um Estado Democrático de Direito em que toda investigação que impacte direitos (no caso, direito à intimidade, privacidade e proteção de dados pessoais) deve partir de uma suspeita fundada. Enquanto o reconhecimento facial estabelece a vigilância a priori de todos os cidadãos de maneira geral, onipresente e remota, isto é, sem controle e percepção do cidadão.

Ademais, a ferramenta em questão não observa os princípios da transparência e da não-discriminação, uma vez que não há indícios dos cuidados que foram adotados para sua utilização e nas bases de dados implementadas para treinamento dos algoritmos. Ainda que as atividades de segurança pública estejam excluídas da aplicação da Lei Geral de Proteção de Dados, precisam estar em adequação aos seus princípios e a própria Lei prevê a necessidade de uma lei regulamentadora sobre a questão. De fato, internacionalmente tem havido iniciativas de suspender o uso dessas tecnologias até uma regulação adequada existir⁹. Como iremos argumentar mais a frente nas recomendações, uma ferramenta dessa natureza, com tantos potenciais de impacto, deveria ter sua implementação sempre precedida de um relatório de impacto.

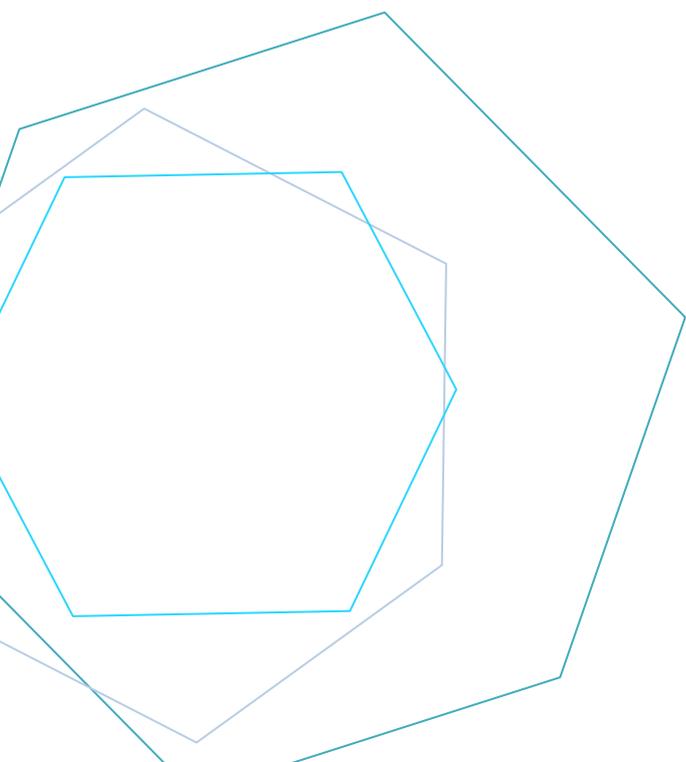
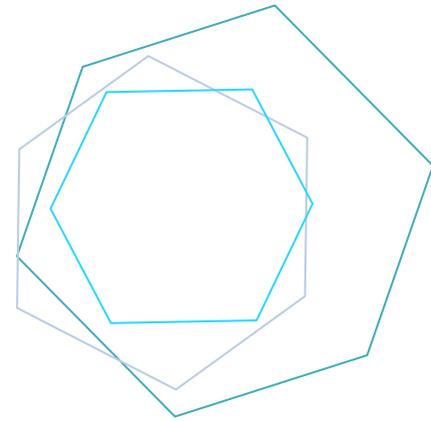
⁹ O caso mais famoso de suspensão do uso de tecnologias de reconhecimento facial é o da cidade de San Francisco, em que os legisladores votaram pela proibição de seu uso junto aos seus habitantes. Em <https://www.bbc.com/news/technology-48276660>



2. Processamento de linguagem natural

Ferramentas de processamento de linguagem natural para estimar algum risco relacionado ao campo penal podem impactar negativamente pessoas de grupos sociodemográficos marginalizados. Por exemplo, a ferramenta Localizador de Evidências Digitais, também usada pelo Departamento de Polícia Federal, que utiliza processamento de linguagem natural para realizar estimativas de risco (incluindo detecção de fraudes).

O algoritmo pode facilmente se tornar enviesado contra pessoas de baixa renda que não possuem um bom domínio da norma culta da língua portuguesa ou que usam termos informais/abreviações/gírias. Este, claro, não é um problema restrito à segurança pública, e viés contra grupos minorizados pode ocorrer com o uso de IA de maneira geral, tendo suas consequências negativas potencializadas no caso da segurança pública, já que muitas vezes resulta em privação de liberdade e mesmo ameaça à vida.



RECOMENDAÇÕES DE GOVERNANÇA

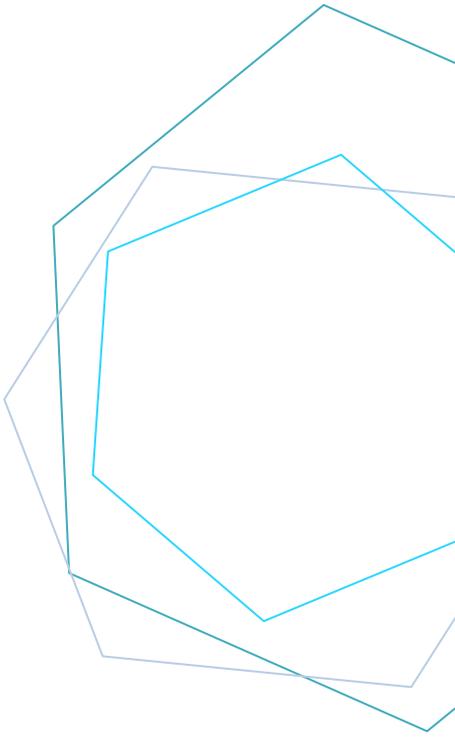
Levando em conta os impactos negativos a direitos apontados na análise acima, apresentamos quatro recomendações de governança, sintetizando as principais preocupações apontadas. A fim de se mitigar estes riscos, sugestões práticas foram propostas a serem consideradas no uso e implementação de ferramentas de IA pelo setor público:

1. Bases de dados representativas e apropriadas para o contexto

O objetivo é prevenir e eliminar/mitigar os vieses dos algoritmos e dos dados de treinamento que poderiam reforçar situações de violência estrutural (como racismo, machismo, LGBTQIfobia, entre outros) em serviços públicos e investigações do aparato de segurança. A estruturação dos bancos de dados de treinamento são questões centrais da análise, não só porque podem impactar direitos relacionados à privacidade e proteção de dados, mas também porque podem ser determinantes para formar ou não vieses potencialmente discriminatórios.

De acordo com as contribuições do InternetLab para a estratégia nacional de Inteligência Artificial, realizada pelo Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), existem maneiras práticas de mitigar tais discriminações – por mais que não possam ser totalmente

17



RECOMENDAÇÕES DE
GOVERNANÇA

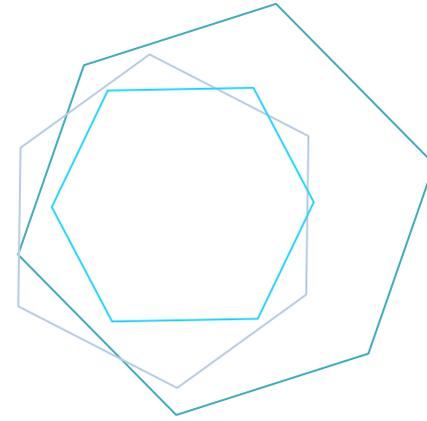
USO DE INTELIGÊNCIA
ARTIFICIAL PELO
PODER PÚBLICO

afastadas. São diversos os tipos de vieses que, por sua vez, podem possuir diferentes origens. Eles podem ocorrer durante as etapas de coleta, limpeza e tratamento dos dados e/ou teste do modelo gerado¹⁰.

Durante a coleta é importante verificar se os dados escolhidos para o conjunto de treinamento do modelo representam bem todos os grupos e populações que serão afetadas por ele ou ainda, se contém alguma discriminação na sua criação que pode vir a reproduzir algum tratamento ou padrão discriminatório. Durante a etapa de limpeza e tratamento e teste do modelo, as crenças pré-existentes dos desenvolvedores do sistema podem contribuir para reproduzir preconceitos existentes. Neste caso, ainda de acordo com o InternetLab, *a principal solução para evitar este tipo de viés é ter diversidade na(s) equipe(s) do projeto.*

Recomendamos, também, que os instrumentos administrativos e contratuais que venham a disciplinar o desenvolvimento e implementação de sistemas de IA no setor público incluam dispositivos/cláusulas com obrigações expressas de utilização de bases de dados representativas das populações presumivelmente afetadas, além da exigência da produção prévia de relatórios de impacto. E que os dados escolhidos para o conjunto de treinamento do modelo devam estar disponíveis para auditoria externa.

¹⁰<https://www.internetlab.org.br/pt/privacidade-e-vigilancia/as-contribuicoes-do-internetlab-para-a-estrategia-nacional-de-inteligencia-artificial/>



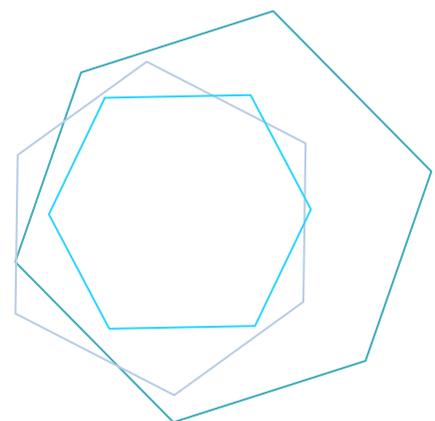
2. Necessidade de supervisão humana como salvaguarda para a revisão de decisões automatizadas

Em relação aos processos de tomada de decisão no poder público com uso de IA, a Lei n. 13.709/2018, chamada de Lei Geral de Proteção de Dados no Brasil (LGPD), prevê em seu artigo 20 que o titular dos dados tem direito a solicitar a revisão de decisões tomadas **unicamente** com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.

O artigo 20, §1º, estabelece que o controlador dos dados deve fornecer as informações de forma clara e adequada a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.

No entanto, percebe-se que a LGPD impõe a condicionante de que a decisão deve ser tomada unicamente com base em tratamento automatizado para obter o direito de revisão. Esta condicionante é um desafio para garantir a salvaguarda da obrigatoriedade da revisão humana das decisões automatizadas, na medida em que, se interpretada de maneira excessivamente restritiva, pode inviabilizar o direito do indivíduo controlar como seus dados são utilizados e os impactos na sua vida.

O mapeamento indica que o resultado obtido por meio de ferramentas de decisões automatizadas é utilizado para influenciar a tomada de decisão governamental. Não sendo

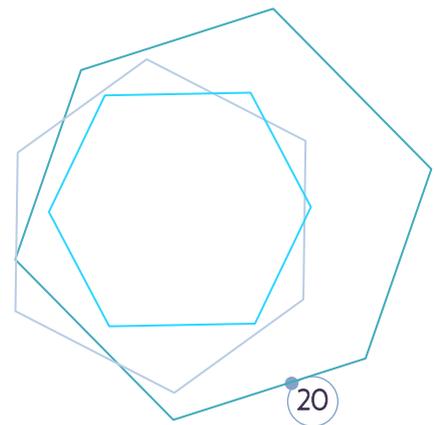


possível, no entanto, afirmar que determinada ação tenha sido tomada, exclusivamente, de forma automatizada pela IA, já que, por haver a interação entre os sistemas e pessoas humanas, é difícil prever o grau de apoio, ou o nível de automatização da ferramenta.

A regulamentação europeia de proteção de dados -- General Data Protection Regulation (GDPR) -- traz a garantia expressa de se contestar a decisão tomada sem supervisão humana. De acordo com tal normativa, os titulares de dados devem ser informados sobre a tomada de decisões automatizadas, além de estarem a par de qual é a lógica aplicada em tais casos, assim como da importância e das consequências dessa tomada de decisão previstas para o titular de dados. A GDPR também determina que ninguém pode ser sujeito a decisões baseadas tão somente em processos automatizados que produzam efeitos legais para o titular de dados ou significativamente o afetem, garantido que os titulares possam solicitar intervenção humana e contestar a decisão.

Recomendamos, assim, que a Autoridade Nacional de Proteção de Dados (ANPD), órgão responsável por zelar, implementar e fiscalizar o cumprimento da LGPD em todo o território nacional, deve interpretar a condicionante do artigo 20 de forma expansiva, já que uma interpretação restritiva do "unicamente" cria um cenário em que o direito à revisão quase nunca poderá ser reivindicado.

Essa interpretação expansiva possui, como fundamento, tanto a principiologia da própria LGPD, quanto o fato de ser o poder público tratando os dados. Ademais, citamos resultados do debate internacional, como a GDPR acima mencionada, e os Princípios de Santa Clara sobre Transparência e Accountability



em Moderação de Conteúdo¹¹, instrumentos que reforçam a necessidade da revisão por pessoa humana como parâmetro mínimo para a proteção dos direitos dos titulares no contexto de decisões automatizadas.

É importante destacar que, durante o processo de aprovação da LGPD, houve o veto ao parágrafo 3º do art. 20, que determinava o direito de revisão como um direito de revisão humana, ou seja, feito por uma pessoa natural. A salvaguarda da revisão humana, portanto, não é garantida pela legislação brasileira atual, que trata apenas de "revisão de decisões tomadas" – não sendo claro qual o tipo de revisão.

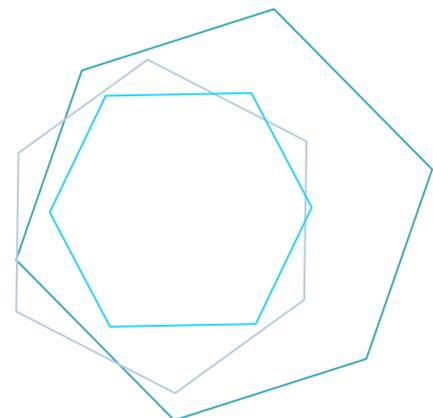
Assim, a legislação deveria refletir tais proteções legais de ampla garantia ao direito de revisão das decisões algorítmicas com revisões humanas. Isso poderia ser feito por regulamentação da LGPD ou ainda por criação de norma específica.

3. Efetiva proteção dos dados pessoais do cidadão

O emprego de tecnologias de IA, em geral, demanda o processamento de grande quantidade de dados para o treinamento do modelo, impulsionando a criação e/ou a disponibilização de bancos de dados pessoais massivos - o que pode ocasionar danos à privacidade do cidadão.

Nesse contexto, existem as seguintes diretrizes trazidas pela

¹¹ Santa Clara Principles on transparency and accountability in content moderation. Disponível em: <https://santaclaraprinciples.org>



LGPD:

(i) O art. 6º dispõe que as atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:

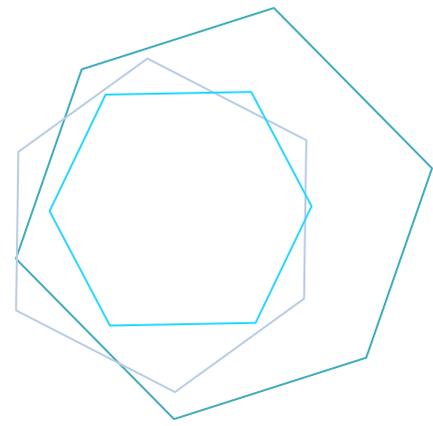
(...)

X - responsabilização e prestação de contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.

(ii) O art. 23 prevê que o tratamento de dados pessoais pelo poder público deverá atender sua finalidade pública. Nesse sentido, a LGPD estabelece como condição para tanto, em seu artigo 23, inciso I, que "*sejam informadas as hipóteses em que, no exercício de suas competências, realizam o tratamento de dados pessoais, fornecendo informações claras e atualizadas sobre a previsão legal, a finalidade, os procedimentos e as práticas utilizadas para a execução dessas atividades, em veículos de fácil acesso, preferencialmente em seus sítios eletrônicos*". Ademais, é fundamental que a observância do art. 6 da LGPD e, em caso de análise, o respeito intransigente aos princípios da finalidade, necessidade, transparência, segurança e não-discriminação.

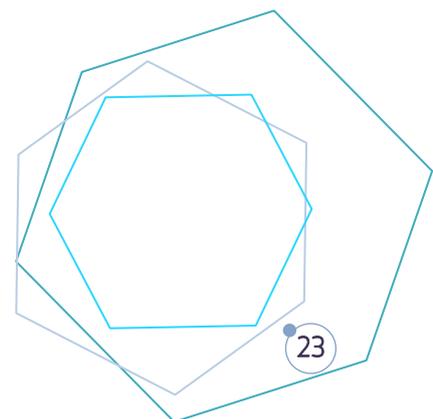
Ou seja, a coleta dos dados deve ser feita estritamente para prestação e melhoria do serviço proposto - com uma finalidade adequada, bem definida e com critérios submetidos ao rigor dos direitos fundamentais.

Neste sentido, é necessário que as pessoas sejam informadas com transparência sobre a finalidade da coleta de seus dados



peçoais e sobre como estão sendo utilizados. É importante entender como foram armazenados os dados para a criação da ferramenta usada por determinado órgão governamental, caso contrário, este serviço estará em desconformidade com os parâmetros estabelecidos pela LGPD.

A ANPD deve fiscalizar o tratamento de dados pessoais pela administração pública, bem como se houve o uso de ferramentas de IA de forma compatível com as finalidades informadas ao titular de dados e com as expectativas legítimas de privacidade e proteção de dados – o que estabelecerá limites ao processamento de dados do algorítmico.



4. Transparência e explicabilidade dos sistemas

Mecanismos de *accountability* e transparência são necessários para que a sociedade civil possa exercer controle social na utilização de ferramentas de inteligência artificial pelo poder público e minimizar os riscos de violações a direitos.

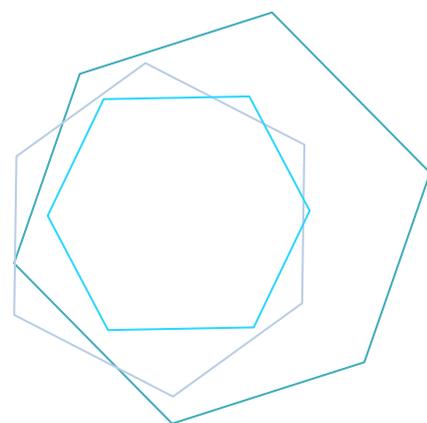
Uma maior transparência do poder público em relação ao uso de algoritmos e tomada de decisão automatizada é importante até mesmo para se avaliar a eficiência dessas ferramentas. Mesmo que não existam ameaças a direitos fundamentais, deve-se avaliar como elas estão sendo usadas, se há um ganho, de fato, na sua utilização, ou ainda, se a ferramenta não está ocasionando prejuízos, como acentuar diferenças sociais e a opressão a grupos marginalizados.

O que significa transparência e *accountability* quando falamos em algoritmos públicos? Fundamentalmente, é preciso garantir

acesso a informações públicas relativas aos algoritmos usados pelo setor público.

Um primeiro caminho é a adoção de algoritmos de IA com códigos abertos. Os códigos devem ser acompanhados de descrições de funcionamento do algoritmo, do banco de dados utilizado para treinamento e, se possível, o próprio banco de dados utilizado ou uma versão anonimizada¹². Ademais, devem existir mecanismos de transparência ativa e o cidadão deve poder solicitar via e-SIC (Sistema Eletrônico do Serviço de Informações ao Cidadão) a relação de quais sistemas algorítmicos o setor público está usando, onde e como.

Quando houver restrição de acesso por razões de sigilo (como sigilo comercial), há dois mecanismos de transparência e *accountability* que devem ser debatidos para escolher o melhor modelo. De maneira mais simples, o poder público deve informar pelo menos as variáveis de entrada e saída do modelo, bem como o tipo de algoritmo utilizado (regressão, redes neurais, árvores de decisão etc.). Em outras palavras, é necessário que as etapas do processo de aprendizado de máquina que resultaram em ação por parte da IA sejam rastreáveis e que as variáveis que pesaram na tomada de decisão sejam públicas.

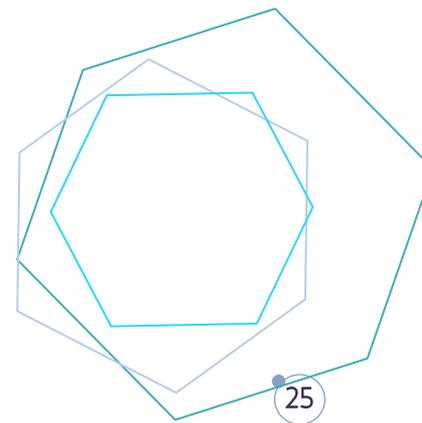


¹² Há muitas possibilidades para ter um banco anonimizado e este documento é agnóstico em relação às soluções que serão adotadas. A título de exemplo, tanto o poder público pode adotar técnicas de privacidade diferencial para anonimizar dados, bem como a criação dos chamados bancos de dados sintéticos. Para saber mais sobre privacidade diferencial e bancos de dados sintéticos, há uma ampla literatura científica sobre essas abordagens, por exemplo, HILTON, Michael. *Differential privacy: a historical survey*. **Cal Poly State University**, 2002; e ABOWD, John M.; LANE, Julia. *New approaches to confidentiality protection: Synthetic data, remote access and research data centers*. In: **International workshop on privacy in statistical databases**. Springer, Berlin, Heidelberg, 2004. p. 282-289, respectivamente.

Outra possibilidade para *accountability* é a realização periódica de auditorias do funcionamento do algoritmo, por especialistas externos à empresa ou órgão público, para averiguar se o algoritmo é eficaz em relação ao seu propósito e se ocasiona algum impacto negativo a direitos fundamentais.

Em terceiro lugar, o governo deve garantir a explicabilidade dos algoritmos, isto é, permitir ao cidadão entender o funcionamento do algoritmo, de que maneira determinada decisão foi tomada, seu propósito e justificativa, além de quais dados são utilizados no processamento. Importante notar que a transparência efetiva nesses casos não se limita à explicabilidade individual, mas deve se centrar na construção de um arcabouço robusto de controle, fiscalização e auditoria desses sistemas pelas autoridades competentes, como o Ministério Público e a ANPD. É preciso ter em mente que nem sempre o cidadão comum terá capacidade de averiguar por si só os resultados da IA, mesmo com bastante transparência. Nesse sentido, produzir a *accountability* horizontal é tão importante quanto a vertical.

Ainda no intuito de se garantir a explicabilidade, sugerimos a obrigatoriedade de elaboração e publicização de Relatório de Impacto Algorítmico prévio à operação, para todos os sistemas de IA que causem impacto no exercício de direitos, que possam prejudicar o cidadão e que envolvam dados sensíveis, especialmente dados genéticos. De acordo com a contribuição à consulta pública para elaboração da Estratégia Brasileira de Inteligência Artificial, enviada pelo Instituto Brasileiro de Defesa do Consumidor (IDEC)¹³ "A elaboração e publicação de



¹³ https://idec.org.br/sites/default/files/sintese_idec_-_estrategia_br_ia.pdf

relatórios prévios também contribui para maior transparência do sistema de IA e compreensão pela sociedade dos riscos existentes, dando suporte para as partes afetadas, para o governo e para o público em geral avaliar quais riscos são ou não aceitáveis, bem como para realizar reclamações e denúncias. Além de ser direito da população saber como os sistemas de IA impactam suas vidas."

Por fim, a avaliação da sociedade civil apontou que sistemas de IA apresentam aplicações com maiores riscos em atividades relacionadas à segurança pública, com consequências sabidamente graves como a privação da liberdade de um indivíduo ou mesmo ameaça à vida. Com isso, medidas de controle público mais consistentes devem ser consideradas com sistemas de IA desenvolvidos para uso para fins de segurança pública e, particularmente, para prevenir o racismo das ferramentas. Na ausência de legislação específica as regulando, o uso de tais tecnologias deve ser vedado pelas autoridades.

ANEXO 01

Catálogo de ferramentas de IA usadas pelo setor público brasileiro

27

ÓRGÃO PÚBLICO

ANATEL

FERRAMENTA

Com base no processamento de linguagem natural, identifica padrões de comportamento do consumidor, com base no registro de reclamações de usuários no sistema da Anatel e provê informações para análise.

IMPACTO NEGATIVO A DIREITOS

O risco está relacionado à existência de outros meios de colocação de reclamação. É necessário que a ferramenta esteja preparada para lidar com as diferentes linguagens, modos de se expressar e de lidar com os temas.

ÓRGÃO PÚBLICO

ANTT

FERRAMENTA

Análise do corpo textual dos tweets

IMPACTO NEGATIVO A DIREITOS

Existe o risco de a ferramenta contribuir para uma personalização e previsão de comportamento futuro do usuário, por meio de rastros de suas preferências manifestadas online. Tal impacto pode restringir o acesso a informações e condicionar o consumo de um determinado produto.

28

ÓRGÃO PÚBLICO

ANTT

FERRAMENTA

Projeção de volume médio diário de tráfego de todas as concessões de rodovias federais

IMPACTO NEGATIVO A DIREITOS

Pode ter impacto no direito de ir e vir das pessoas, já que os critérios a serem utilizados pelo modelo podem levar a reprodução de uma discriminação social já pré-existente, ou mesmo criar elementos. Por exemplo: a ferramenta pode sugerir ações no sentido de organizar o fluxo de trânsito para lugares com mais ou menos veículos, ou com mais ou menos acidentes. Uma localização com uma maior concentração de pessoas vulneráveis pode ter mais acidentes não necessariamente pelo número absoluto de carros, mas sim por ter mais veículos em mau estado de conservação.

ÓRGÃO PÚBLICO

ANVISA

FERRAMENTA

Análise das contribuições às Consultas Públicas publicadas, por meio de agrupamento de informações de texto

IMPACTO NEGATIVO A DIREITOS

A ferramenta pode impactar direitos de acesso a informações e de expressão, a depender do modo como o algoritmo foi treinado e quais elementos efetivamente a ferramenta vai levar em consideração para agrupar e dar prioridade a contribuições. Tanto um "bom" quanto um mau funcionamento defeituoso podem impactar na escuta que o órgão faz das contribuições da sociedade civil em temas de alto impacto como a liberação de alimentos geneticamente modificados (OGMs) e agrotóxicos, impactando assim no direito da sociedade de incidir nas políticas públicas.

ÓRGÃO PÚBLICO

Banco do Brasil

FERRAMENTA

Chatbot para dar atendimento e respostas a dúvidas comuns sobre produtos e serviços

IMPACTO NEGATIVO A DIREITOS

O receio com ferramentas de chatbot de modo geral é o acesso a serviço por pessoas que sejam analfabetas, migrantes ou que não tenham acesso estável à internet. O chatbot é a porta de entrada para a obtenção de serviços e o exercício de uma série de direitos. Deve existir uma série de cuidados, desde o ponto de vista da linguagem utilizada para ser compreensível e acessível, até se não há um favorecimento em nenhum dos pontos. Além do mais, se for um canal principal de comunicação deve ser pensado para ser inclusivo, pensando nos diferentes elementos de conectividade (dispositivo, banda...).

ÓRGÃO PÚBLICO

Banco do Brasil

FERRAMENTA

Ferramenta que reconhece se um rosto de uma selfie é o mesmo rosto de um documento pessoal (CNH e RG) e classifica se um documento é uma CNH ou RG, na abertura de conta fácil (conta digital)

IMPACTO NEGATIVO A DIREITOS

Existem riscos relacionados ao treinamento da ferramenta e de precisão do algoritmo, com consequências possíveis de exclusão. Os cuidados necessários se relacionam à proteção de dados e à segurança da informação.

30

ÓRGÃO PÚBLICO

Banco do Brasil

FERRAMENTA

Ferramenta para triagem automatizada de ofícios jurídicos que classifica o tipo de documento e faz a extração de informações como data de emissão, comarca, estado, data de vencimento.

IMPACTO NEGATIVO A DIREITOS

Mesmo em seu bom uso o sistema pode excluir documentos e impactar de maneira negativa usos e costumes locais, impedindo o acesso a serviços públicos.

ÓRGÃO PÚBLICO

Banco do Brasil

FERRAMENTA

Ferramenta que estima a probabilidade de determinada transação eletrônica ser fraudulenta

IMPACTO NEGATIVO A DIREITOS

Deve-se considerar possíveis vieses que podem impactar negativamente pessoas de baixa renda ou pertencentes a minorias políticas. Uma análise mais detida dos proxies para "operação fraudulenta" se faz necessária.

31

ÓRGÃO PÚBLICO

BNDES

FERRAMENTA

Ferramenta para esclarecimento de dúvidas e orientações sobre os programas emergenciais

IMPACTO NEGATIVO A DIREITOS

Qualquer viés no processamento da linguagem poderá impactar negativamente minorias políticas que não possuem um bom domínio da linguagem escrita, dificultando ainda mais seu acesso aos serviços ofertados pelo BNDES. É importante que seja determinada a obrigatoriedade de revisão humana dos outputs gerados pelo chatbot.

ÓRGÃO PÚBLICO

BNDES

FERRAMENTA

Ferramenta faz o diagnóstico de identificação do parceiro de negócio que melhor atenderá a proposta de financiamento baseado na análise do perfil da proposta do BNDES

IMPACTO NEGATIVO A DIREITOS

Mesmo o bom funcionamento desta ferramenta pode gerar exclusão, pois a tendência é focalizar os recursos naqueles elementos que historicamente já tiveram acesso ao Banco.

32

ÓRGÃO PÚBLICO

CADE – Conselho Administrativo de Defesa Econômica

FERRAMENTA

Ferramenta para diagnóstico de probabilidade de ocorrência de acordo entre empresas licitantes.

IMPACTO NEGATIVO A DIREITOS

O modelo pode impactar direitos de consumidores, investidores e de outras empresas. A tendência do histórico deve ser observar acordos maiores o que pode gerar uma distorção. Igualmente, pode existir a concentração de recursos levando a potenciais questões de inclusão e acesso.

ÓRGÃO PÚBLICO

Caixa Econômica Federal

FERRAMENTA

Chatbot para conversa guiada e transações bancárias

IMPACTO NEGATIVO A DIREITOS

A relacionada a chatbots traz como riscos tanto questões de reconhecimento da linguagem, quanto de meios de acesso. A linguagem deve ser inclusiva, pensada para os diferentes contextos e grupos. Podem existir também elementos na arquitetura e no fluxo dos chatbots que podem restringir o exercício de outros direitos. O receio com ferramentas de chatbot de modo geral é o acesso a serviço por pessoas que sejam analfabetas, migrantes ou que tenham acesso instável à internet.

33

ÓRGÃO PÚBLICO

Caixa Econômica Federal

FERRAMENTA

Ferramenta para analisar detecção de fraude, suspeitas de fraude, bloqueios de cadastro, autorização de transações com cartões de débito e crédito.

IMPACTO NEGATIVO A DIREITOS

O uso de modelos de predição de fraude pode ter um impacto profundo na vida das pessoas. O cuidado com os elementos específicos que são analisados é crucial. Muitas vezes modelos preditivos de fraude tomam em consideração fatores que podem não necessariamente ter uma clara relação causal. Nesse sentido, podem ferir em muito as expectativas das pessoas. A ferramenta pode impactar o direito de acesso a serviços financeiros ao não autorizar transações de cartão de débito ou crédito baseada em vieses discriminatórios.

RECOMENDAÇÕES DE
GOVERNANÇA

USO DE INTELIGÊNCIA
ARTIFICIAL PELO
PODER PÚBLICO

ÓRGÃO PÚBLICO

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

FERRAMENTA

Ferramenta para diagnóstico de probabilidade de: i) duas produções científicas serem a mesma; ii) duas pessoas, não identificadas, serem a mesma; iii) duas instituições (pública ou privada, internacional ou nacional) serem a mesma; iv) dois projetos de pesquisa serem os mesmos ou terem os mesmos financiadores

IMPACTO NEGATIVO A DIREITOS

Qualquer viés ou erro de funcionamento no processo de análise de trabalhos científicos poderá impactar negativamente acadêmicos que dependem das métricas da CAPES para exercer certos direitos (ex.: progressão na carreira de docente).

ÓRGÃO PÚBLICO

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

FERRAMENTA

Ferramenta para sugerir recomendações sobre termos de busca para periódicos.

IMPACTO NEGATIVO A DIREITOS

Qualquer viés ou erro de funcionamento no processo de classificação de um periódico poderá impactar negativamente acadêmicos que dependem das métricas da CAPES para exercer certos direitos (ex.: progressão na carreira de docente)

ÓRGÃO PÚBLICO

CGU - Controladoria Geral da União

FERRAMENTA

Ferramenta para sugerir a probabilidade de a prestação de contas de um convênio possuir problemas

IMPACTO NEGATIVO A DIREITOS

É importante que a necessidade de um humano analisar o caso concreto seja mantida, evitando, assim, que qualquer viés ou problema no funcionamento da ferramenta seja sedimentado.

35

ÓRGÃO PÚBLICO

CGU - Controladoria Geral da União

FERRAMENTA

Ferramenta para avaliar a probabilidade de determinado caso apresentar fraude ou irregularidade

IMPACTO NEGATIVO A DIREITOS

Preocupa que eventuais vieses do algoritmo favoreçam ou prejudiquem administradores(as) locais em razão de preferências políticas. A depender de como seja treinado, pode ser inócuo para detectar algumas fraudes enquanto dê falsos-positivos para outra, desviando o foco do trabalho investigativo.

ÓRGÃO PÚBLICO

DPF – Departamento de Polícia Federal

FERRAMENTA

Registro de informações da fase sigilosa das investigações. A ferramenta utiliza técnicas de reconhecimento de imagens, reconhecimento facial e perfil criminal.

IMPACTO NEGATIVO A DIREITOS

Ferramentas de reconhecimento facial podem levar a uma série de vieses a depender do banco de dados usado para seu treinamento. Um viés possível diz respeito ao não reconhecimento de pessoas negras ou então à associação enviesada de pessoas negras à atividades criminosas. Esses vieses podem impactar negativamente o direito ao tratamento igualitário.

A ferramenta pode não sugerir ações, mas certamente pelas técnicas descritas classifica pessoas entre mais ou menos probabilidade de suspeição e isso determina o rumo das investigações.

Outra preocupação é a custódia dos bancos de dados de imagens (já que a biometria facial é um dado pessoal sensível) e vieses em razão da forma como os algoritmos são treinados.

ÓRGÃO PÚBLICO

DPF – Departamento de Polícia Federal

FERRAMENTA

Ferramenta de classificação de imagens que estima o índice de determinada imagem conter nudez.

IMPACTO NEGATIVO A DIREITOS

Ferramentas de reconhecimento facial podem levar a uma série de vieses a depender do banco de dados usado para seu treinamento. Um viés possível diz respeito ao não reconhecimento de pessoas negras ou então à associação enviesada de pessoas negras à atividades criminosas.

O uso de ferramentas para detectar nudez em uma imagem também pode ter impactos distintos a depender do sexo do sujeito em questão. Esses vieses podem impactar negativamente o direito ao tratamento igualitário.

É importante também que a base nacional de arquivos contendo cenas de abuso sexual infantil seja diversificada o suficiente para mitigar o risco desses vieses.

A revisão humana certamente não é feita para todas as imagens. Nesse caso um possível erro de funcionamento impactaria negativamente, podendo deixar de identificar uma violação de direito fundamental da criança

ÓRGÃO PÚBLICO

DPF – Departamento de Polícia Federal

FERRAMENTA

Ferramenta de reconhecimento facial e classificação de imagens que: i) auxilia na identificação de rostos; ii) classifica rostos em grupos de idades, iii) identificar objetos em cenas e, iv) estima a probabilidade de determinada imagem conter nudez.

IMPACTO NEGATIVO A DIREITOS

Sem considerar possíveis vieses, é possível que o funcionamento gere distorções e impactos negativos em direitos. Ferramentas de reconhecimento facial tendem a ter menor acurácia na identificação e categorização de pessoas negras. Isso depende principalmente do banco de dados utilizado e dos algoritmos.

Considerando que a ferramenta foi tratada com a base nacional de arquivos com cenas de abuso sexual infantil, cabe analisar a diversidade de gênero, etnia e raça desta base para evitar possíveis vieses.

Ferramentas de processamento de linguagem natural para estimar algum risco relacionado ao campo penal podem impactar negativamente pessoas marginalizadas. O algoritmo pode facilmente se tornar enviesado contra pessoas de baixa renda que não possuem um bom domínio do português ou que usam termos informais/abreviações/gírias.

ÓRGÃO PÚBLICO

DPF – Departamento de Polícia Federal

FERRAMENTA

Ferramenta que utiliza processamento de linguagem natural para realizar estimativas de risco (incluindo detecção de fraudes). A ferramenta auxilia no reconhecimento de entidades (nomes de pessoas, empresas, endereços, valores, e-mails, números de telefone, etc)

IMPACTO NEGATIVO A DIREITOS

Sistemas de reconhecimento de linguagem natural na segurança pública, quando não analisados para detecção de possíveis vieses, podem impactar negativamente o princípio da presunção de inocência.

Além disso, ainda que as atividades de segurança pública estejam excluídas da aplicação da Lei Geral de Proteção de Dados, precisam estar em adequação aos seus princípios. A ferramenta em questão não observa os princípios da transparência e da não-discriminação, uma vez que não há indícios dos cuidados que foram adotados para sua utilização.

Trata-se de um impacto reproduzido do ambiente social, por meio de bases de dados possivelmente enviesadas, que acabam sendo acentuados pela utilização da ferramenta. O fato de o sistema funcionar melhor com textos que seguem a norma culta da língua pode levar a falsos positivos ou falsos negativos, prejudicando pessoas com diferentes níveis de escolaridade e a ausência de transparência sobre os critérios aplicados para evitar possíveis vieses, podem acentuar o impacto negativo da ferramenta.

Ferramentas de processamento de linguagem natural para estimar algum risco relacionado ao campo penal podem impactar negativamente pessoas marginalizadas. O algoritmo pode facilmente se tornar enviesado contra pessoas de baixa renda que não possuem um bom domínio do português ou que usam termos informais/abreviações/gírias.

ÓRGÃO PÚBLICO

EBC – Empresa Brasil de Comunicação S.A.

FERRAMENTA

Ferramenta de classificação de imagens para detecção de reprodução de imagens da TV Brasil na programação das emissoras parceiras.

IMPACTO NEGATIVO A DIREITOS

Sem considerações

ÓRGÃO PÚBLICO

EBSERH - HU-UFMA - Hospital Universitário da Universidade Federal do Maranhão

FERRAMENTA

Chatbot para realizar a triagem de pacientes com sintomas de covid-19. Ferramenta classifica o risco de infecção por coronavírus dos pacientes.

IMPACTO NEGATIVO A DIREITOS

É importante considerar que vieses e problemas de triagem gerados pelo chatbot (ex.: não reconhecimento de uma combinação específica de sintomas como um indicativo de infecção por coronavírus) podem impactar negativamente o direito à saúde do cidadão.

Ademais, o direito à saúde pode ser afetado, já que a partir de uma sequência pré definida de comandos automatizados que com base em dados pessoais e não pessoais chegam a conclusões que podem impactar significativamente a vida do cidadão, que consegue pedir informações adicionais ou explicações imediatas sobre como foi chegado a esta decisão.

A depender dos critérios utilizados, as triagens e classificação de risco podem ser discriminatórias, favorecendo, como outros casos já demonstraram, pessoas brancas e de classe alta

ÓRGÃO PÚBLICO

Embrapa Agroenergia

FERRAMENTA

Ferramenta para a identificação e classificação de espécimes vegetais a partir de fotografias tiradas com o auxílio de drones, feita em Tensor Flow.

IMPACTO NEGATIVO A DIREITOS

Apesar desta ferramenta ser específica para a classificação de espécies vegetais, a geração de imagens por meio de sobrevoos de drones, quando realizados sobre territórios indígenas e quilombolas demandam a consulta prévia, livre e informada aos moradores, antes da realização das atividades. A realização de projetos ou iniciativas de registros e/ou levantamentos em territórios indígenas e quilombolas sem a consulta prévia aos povos indígenas e populações tradicionais ferem as regulamentações estabelecidas pela OIT 169.

Existe a possibilidade dessa ferramenta impactar de maneira indireta partes da população. Depende não somente dela, mas sim de como ela for distribuída. Há sempre o potencial de erro e de precisão, mas funcionando bem, o seu impacto pode ser no sentido de aumentar o fosso entre os que podem ter acesso a ela e os que não têm.

ÓRGÃO PÚBLICO

Embrapa Gado de Corte

FERRAMENTA

Algoritmo preditivo que informa ao produtor qual a melhor cultivar a ser plantada em áreas de sua fazenda com base em informações agronômicas.

IMPACTO NEGATIVO A DIREITOS

Apesar desta ferramenta ser bem específica para cultivo de pasto para gado, ela levanta a questão do que seria um cultivo recomendado. Esta mesma lógica pode, no futuro, ser transposta para outras culturas e ter um impacto na segurança alimentar de populações, que sofrem por exemplo com a pressão da "produtividade" e são impelidas ao uso de agrotóxicos; da mesma forma, poderiam no futuro se sentirem forçadas a adotarem algoritmos preditivos com base em eficiência/produtividade de longo prazo. Isto depende muito de como o algoritmo foi construído e/ou treinado considerando o que seria um cultivo apropriado.

O viés aqui pode tender explicitamente para a lógica do agronegócio e da monocultura que desconsidera práticas tradicionais e levando em consideração apenas variáveis agronômicas compatíveis com análise de 5 especialistas e obtenção de resultados de curto prazo; outros tipos de saberes e conhecimentos precisam ser considerados se o sistema pretender um uso geral.

No caso específico da pastagem para gado, fica a pergunta se a ferramenta funciona também para indicar qual o melhor cultivar para terrenos recém queimados/desmatados OU se o sistema terá algum tipo de restrição para não fornecer informações para terrenos em situação irregular, isto é, referentes à frente de expansão destruidora do agronegócio sem compromisso com o ambiente (isto é, onde o "Pasto é certo" e onde ele é errado).

Existe a possibilidade dessa ferramenta impactar de maneira indireta partes da população, pois há sempre o potencial de erro e de precisão, mas funcionando bem, o seu impacto pode ser no sentido de aumentar o fosso entre os que podem ter acesso a ela e os que não têm.

ÓRGÃO PÚBLICO

Embrapa Informática Agropecuária

FERRAMENTA

Ferramenta para classificação imagens com doenças de plantas em algumas culturas agrícolas (ferramenta para atender demandas de pesquisa).

IMPACTO NEGATIVO A DIREITOS

Existe a possibilidade dessa ferramenta impactar de maneira indireta partes da população. Depende não somente dela, mas sim de como ela for distribuída. Há sempre o potencial de erro e de precisão, mas funcionando bem, o seu impacto pode ser no sentido de aumentar o fosso entre os que podem ter acesso a ela e os que não têm.

43

ÓRGÃO PÚBLICO

Embrapa Informática Agropecuária

FERRAMENTA

Ferramenta de classificação de imagens para estimativas de produção/contagem de frutos (fruticultura de precisão) (em desenvolvimento).

IMPACTO NEGATIVO A DIREITOS

Existe a possibilidade dessa ferramenta impactar de maneira indireta partes da população. Depende não somente dela, mas sim de como ela for distribuída. Há sempre o potencial de erro e de precisão, mas funcionando bem, o seu impacto pode ser no sentido de aumentar o fosso entre os que podem ter acesso a ela e os que não têm.

RECOMENDAÇÕES DE
GOVERNANÇA

USO DE INTELIGÊNCIA
ARTIFICIAL PELO
PODER PÚBLICO

ÓRGÃO PÚBLICO

Embrapa Trigo

FERRAMENTA

Ferramenta que classifica a probabilidade de identificar doenças em folhas de trigo (em desenvolvimento)

IMPACTO NEGATIVO A DIREITOS

Existe a possibilidade dessa ferramenta impactar de maneira indireta partes da população. Depende não somente dela, mas sim de como ela for distribuída. Há sempre o potencial de erro e de precisão, mas funcionando bem, o seu impacto pode ser no sentido de aumentar o fosso entre os que podem ter acesso a ela e os que não têm.

ÓRGÃO PÚBLICO

FURG – Fundação Universidade Federal do Rio Grande

FERRAMENTA

Chatbot que atende público específico para reconhecer as intenções no uso do AVA (ambiente virtual de aprendizagem)

IMPACTO NEGATIVO A DIREITOS

Por já ser direcionado a um público específico é importante entender se aquelas pessoas foram informadas com transparência sobre a finalidade da coleta dos dados e como estão sendo utilizados aqueles dados.

Ou seja, é importante entender como foram armazenados os dados para a criação deste chatbot para o público específico e se eles foram informados deste armazenamento, caso contrário, este serviço estará cometendo um abuso quanto ao uso de informações pessoais dos seus usuários.

Além de possivelmente privar que um certo grupo de pessoas tenha acesso a estas informações (já que seleciona um perfil específico) e influenciar as pessoas com acesso a esta informação a escolhas específicas. O serviço também não passa por uma revisão de humana, o que impossibilita o direito de explicação para o sujeito receber informações suficientes e inteligíveis para entender a lógica e critério utilizados para tratar seus dados.

Se o output do chatbot representar apenas uma sugestão, a falta de revisão humana não deve ser um problema. Mas caso o output passe a representar uma condição de acesso ao AVA, a falta de revisão humana representa uma violação ao direito garantido pelo art. 20 da LGPD.

ÓRGÃO PÚBLICO

HFA – Hospital das Forças Armadas

FERRAMENTA

Ferramenta para monitoramento do estado de saúde de pacientes. Triagem virtual para classificar risco de Covid-19

IMPACTO NEGATIVO A DIREITOS

A depender dos critérios utilizados, as triagens e classificação de risco podem ser discriminatórias, favorecendo, como outros casos já demonstraram, pessoas brancas e de classe alta. É imprescindível que exista revisão humana nesse caso. Qualquer triagem enviesada ou que apresente problemas pode impactar negativamente o direito à saúde do cidadão. Ademais, segundo o art. 20 da LGPD, qualquer decisão automatizada baseada no tratamento de dados pessoais confere ao titular o direito de solicitar sua revisão por um humano.

ÓRGÃO PÚBLICO

INSS – Instituto Nacional do Seguro Social

FERRAMENTA

Ferramenta de análise preditiva de padrões na concessão de benefícios buscando indícios de irregularidades.

A análise realizada pela solução consiste em identificar desvios padrões de comportamento esperado na análise e concessão de determinado benefício. A ferramenta apenas identifica indícios de fraude e respectivo volume dentro do conjunto de dados analisado.

IMPACTO NEGATIVO A DIREITOS

Sem considerações

ÓRGÃO PÚBLICO

UFRN – Universidade Federal do Rio Grande do Norte

FERRAMENTA

Ferramenta para otimizar o atendimento dos usuários pela ouvidoria e o atendimento quanto aos pedidos via Lei de Acesso à Informação.

IMPACTO NEGATIVO A DIREITOS

O possível viés da ferramenta (com maior precisão em relação aos alunos, em detrimento dos servidores) pode impactar negativamente o direito de acesso à informação.

47

ÓRGÃO PÚBLICO

UFSM – Universidade Federal de Santa Maria

FERRAMENTA

Ferramenta para otimizar o atendimento dos usuários pela ouvidoria e o atendimento quanto aos pedidos via Lei de Acesso à Informação.

IMPACTO NEGATIVO A DIREITOS

Se a ferramenta for utilizada para tomada de ações que penalizam aqueles classificados com alta probabilidade de abandono (como negação de benefício de auxílios estudantis), pode impactar negativamente no direito à educação.

Aspectos socioeconômicos devem ser incluídos no modelo, estudantes com dificuldades financeiras familiares, mas com bom desempenho acadêmico também podem abandonar o ensino. Se ações são sugeridas, esse conjunto de pessoas pode estar deixando de ter acesso.

Se as decisões, com o tempo, forem baseadas apenas nos outputs do algoritmo, segundo o art. 20 da LGPD o titular dos dados pessoais (nesse caso, o acadêmico) terá direito a solicitar a revisão humana.

ÓRGÃO PÚBLICO

Tribunal Superior do Trabalho

FERRAMENTA

Ferramenta para classificação de processos e previsões sobre a tramitação do processo nos Gabinetes

IMPACTO NEGATIVO A DIREITOS

A ferramenta pode trazer prejuízos ao acesso à justiça, ao devido processo legal e ao dever legal de motivação das decisões judiciais caso automatize decisões que deveriam ser fruto de uma escolha humana. Ainda que, num primeiro momento, a ferramenta seja apenas sugestiva, há que se analisar com muita atenção e transparência a forma de construção do algoritmo e a maneira como é utilizado: não se pode tratar o processo decisório como mera questão de gestão.

As sugestões realizadas pela ferramenta podem influenciar as decisões dos Ministros; assim, se incorporarem decisões passadas corroboram para não evolução da jurisprudência, mesmo com bom funcionamento; ou, se pelo mau funcionamento, indicarem soluções inadequadas ao caso, podem diminuir a qualidade da prestação jurisdicional. Podem impactar, assim, o acesso à justiça, que é um guarda-chuva para todos os direitos trabalhistas.

É importante que o output do algoritmo permaneça como uma sugestão. Qualquer viés ou problema no seu funcionamento poderá impactar negativamente o direito a uma decisão judicial fundamentada ou então de acesso à justiça.

ÓRGÃO PÚBLICO

Supremo Tribunal Federal (STF)

FERRAMENTA

Ferramenta para classificação de processos judiciais em temas de repercussão geral, que objetiva simplificar o reconhecimento de padrões em textos jurídicos apresentados perante o Supremo Tribunal Federal

IMPACTO NEGATIVO A DIREITOS

O algoritmo pode produzir assimetria e tratamentos desiguais: alguém que conheça seu funcionamento terá maiores condições de conseguir que seu recurso seja conhecido ao aplicar engenharia reversa, uma vez que poderá utilizar as palavras-chave vinculadas ao que o modelo entende por "mais provável de ser conhecido" e, com isso, gerar manipulação. É necessário que seja dada ampla transparência desse processo.

A classificação de processos em temas de repercussão geral é um importante aspecto da eficiência jurisdicional, ainda mais considerando o importante papel do STF na estrutura do poder judiciário. Qualquer viés ou problema no funcionamento desse algoritmo poderia impactar negativamente o direito a uma prestação jurisdicional eficiente e célere.

ÓRGÃO PÚBLICO

Superior Tribunal de Justiça (STJ)

FERRAMENTA

Ferramenta: produz um exame automatizado de cada recurso encaminhado ao STJ e decisões prévias do processo, recomenda fontes normativas e precedentes jurídicos, e fornece uma recomendação de ação (a decisão final sempre será realizada pelo Ministro do STJ).

IMPACTO NEGATIVO A DIREITOS

O algoritmo, ao aprender a partir de decisões já tomadas, pode reproduzir preconceitos sociais embutidos nos tribunais/tomadores de decisões. Apesar de não tomar decisões por si, e sim apenas “aumentar a eficiência e velocidade de avaliação judicial”, o fato de ser utilizada pode gerar a falsa sensação de imparcialidade do estudo de jurisprudência realizado por ela.

Idem os casos acima. Preocupa a automatização do processo decisório e o desconhecimento quanto à aplicação da ferramenta: teremos dados sobre sua utilização? Sabemos o índice de “adesão” do Ministro à sugestão do sistema? Como essas bases são treinadas?

Agravamento por um possível mau uso: opacidade do processo decisório.

É importante que o output do algoritmo permaneça como uma sugestão. Qualquer viés ou problema no seu funcionamento poderá impactar negativamente o direito a uma decisão judicial fundamentada.

ÓRGÃO PÚBLICO

Tribunal de Contas da União (TCU)

FERRAMENTA

Ferramenta de classificação textual de documentos em PDF associados a danos ao erário público, inseridos no sistema de gestão de Tomadas de Contas Especiais (e-TCE) do TCU.

IMPACTO NEGATIVO A DIREITOS

Como apontado, há possibilidade de reforço de desigualdades entre pessoas que tenham maior conhecimento da ferramenta em detrimento das demais. Nesse caso, é importante que o output do algoritmo permaneça apenas como uma sugestão. Um possível viés ou problema no seu funcionamento poderia fazer com que o TCU deixasse de monitorar documentos sensíveis.

51

ÓRGÃO PÚBLICO

Tribunal de Contas da União (TCU)

FERRAMENTA

Ferramenta para auxiliar na correção de acórdãos do TCU, sem erros materiais investigados, tem como output um conjunto de alertas de incorreções materiais (exemplo: CPF inválido)

IMPACTO NEGATIVO A DIREITOS

Um possível viés ou problema no funcionamento desse algoritmo pode tornar a prestação da justiça menos célere, impactando negativamente o direito a um processo jurisdicional célere e eficiente.

ÓRGÃO PÚBLICO

Tribunal de Contas da União (TCU)

FERRAMENTA

Chatbot. Plataforma facilitadora para acesso às soluções públicas do TCU

IMPACTO NEGATIVO A DIREITOS

Um possível viés ou problema no funcionamento desse algoritmo poderia impactar negativamente o direito de acesso à informação.

52

ÓRGÃO PÚBLICO

Tribunal de Contas da União (TCU)

FERRAMENTA

Ferramenta: Instrução assistida de pareceres sobre processos judiciais do Tribunal de Contas da União (TCU)

IMPACTO NEGATIVO A DIREITOS

A ferramenta, ao ser treinada a partir de pareceres emitidos previamente, pode reproduzir tópicos ou decisões que afetem negativamente direitos, especialmente de minorias que costumam ser negligenciadas ou não acolhidas pelas decisões do sistema judiciário.

Um possível viés nesse caso pode impactar negativamente o direito de acesso à justiça e ao devido processo legal. É importante que qualquer fase processual, ainda que apenas a confecção de um parecer jurídico, leve em consideração todas as informações pertinentes ao caso.

ÓRGÃO PÚBLICO

Tribunal de Contas da União (TCU)

FERRAMENTA

Extração de deliberações dos acórdãos proferidos pelo TCU.

IMPACTO NEGATIVO A DIREITOS

O possível impacto a algum direito dependerá da finalidade para a qual a classificação seria utilizada. Sendo apenas uma sugestão, não parece que há maiores problemas.

53

ÓRGÃO PÚBLICO

Tribunal de Contas da União (TCU)

FERRAMENTA

Classificação de textos do Tribunal de Contas da União (TCU)

IMPACTO NEGATIVO A DIREITOS

Ferramentas de classificação de demandas podem impactar no direito de acesso à justiça, é importante, portanto, que seus outputs sejam considerados apenas sugestões e que exista a revisão humana.

ÓRGÃO PÚBLICO

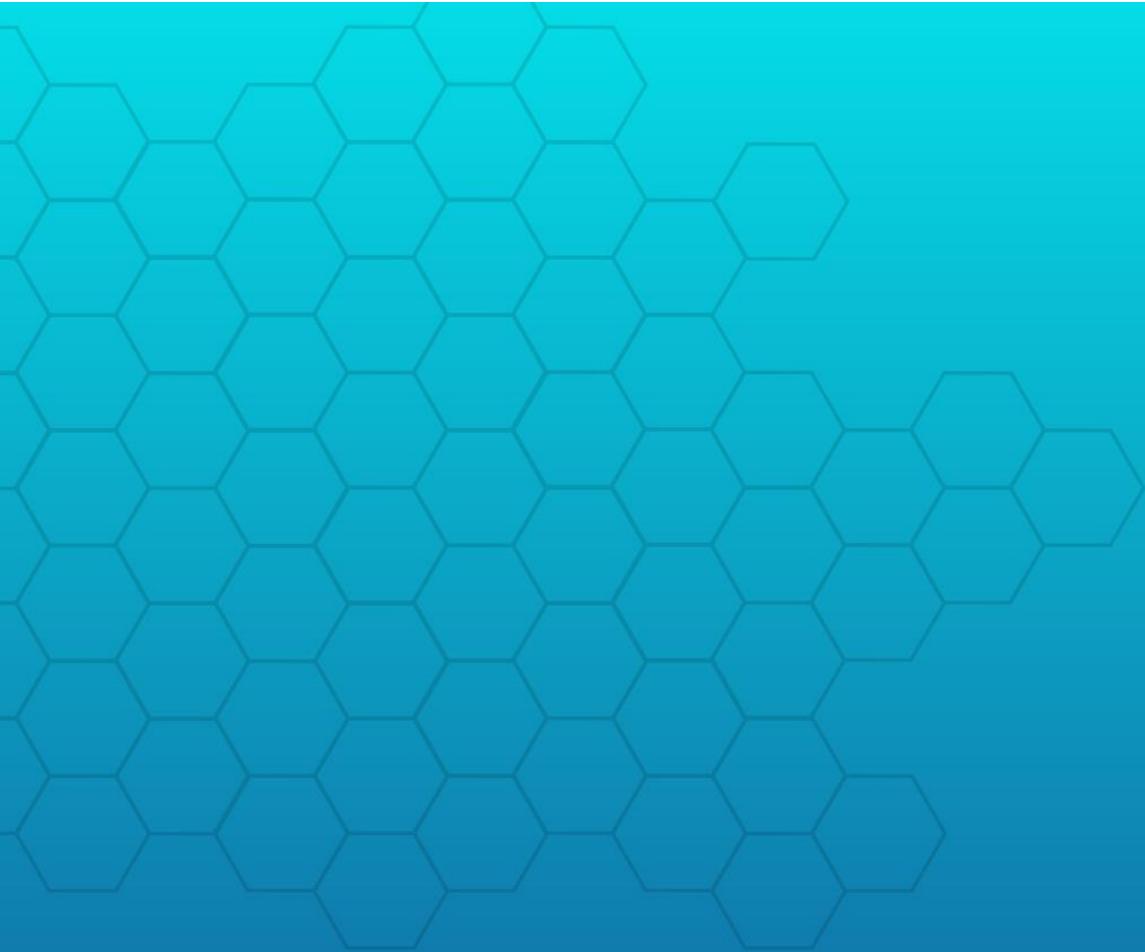
Tribunal de Contas da União (TCU)

FERRAMENTA

Ferramenta de cruzamento de dados e estrutura em grafos para análise de grandes volumes de dados. A ferramenta realiza a extração de relações entre pessoas físicas e jurídicas em processos judiciais e as exibe em grafos para posterior análise.

IMPACTO NEGATIVO A DIREITOS

A falha na revisão de documentos digitalizados pode gerar a seleção arbitrária de dados e, logo, excluir determinados processos judiciais das análises que serão geradas a partir do uso da ferramenta.



TRANSPARÊNCIA
BRASIL

transparencia.org.br

