GOVERNANCE RECOMMENDATIONS

USE OF ARTIFICIAL INTELLIGENCE BY PUBLIC AUTHORITIES

TRANSPARÊNCIA BRASIL

BY

TRANSPARÊNCIA BRASIL

CHIEF EXECUTIVE
Manoel Galdino

CHIEF OPERATING OFFICER
Juliana Sakai

PROJECT COORDINATOR
Tamara Burg

RESEARCH
Jonas Coelho
Tamara Burg

TEXT
Juliana Sakai
Manoel Galdino
Tamara Burg

TEXT FORMATTING
Marina Atoji

www.transparencia.org.br

FUNDING

ICNL
INTERNATIONAL CENTER FOR NOT-FOR-PROFIT LAW

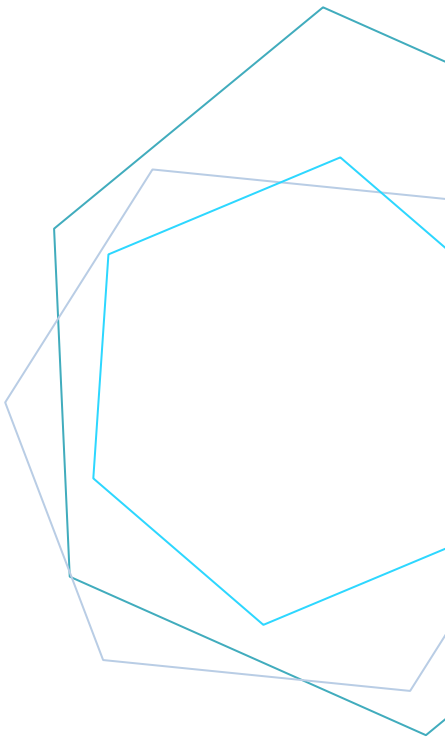
PARTNERSHIP
Northwestern University

CONTRIBUTION
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

SUMMARY

INTRODUCTION

This document presents recommendations by Transparência Brasil for the use and development of technology related to Artificial Intelligence (AI)[1] by the Brazilian public sector. They have been made in collaboration with 12 civil society organizations[2]: Artigo 19 (Article 19); Conectas Direitos Humanos (Conectas Human Rights); Instituto Brasileiro de Defesa do Consumidor (Brazilian Institute for Consumer Protection, IDEC); Instituto de Defesa do Direito de Defesa (Institute for Rights of Defense, IDDD); Instituto de Estudos da Religião (Religion Studies Institute, ISER); Instituto de Referência em Internet e Sociedade (Institute for Research on Internet and Society, IRIS); Instituto de Tecnologia e Sociedade (Institute for Technology and Society, ITS); Instituto Igarapé (Igarapé Institute); Instituto Socioambiental (Socioenvironmental Institute, ISA); Minas Programam (Girls who Code); Mulheres Negras Decidem (Black Women Decide) and PretaLab[3]



[1] OECD defines artificial intelligence systems as "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

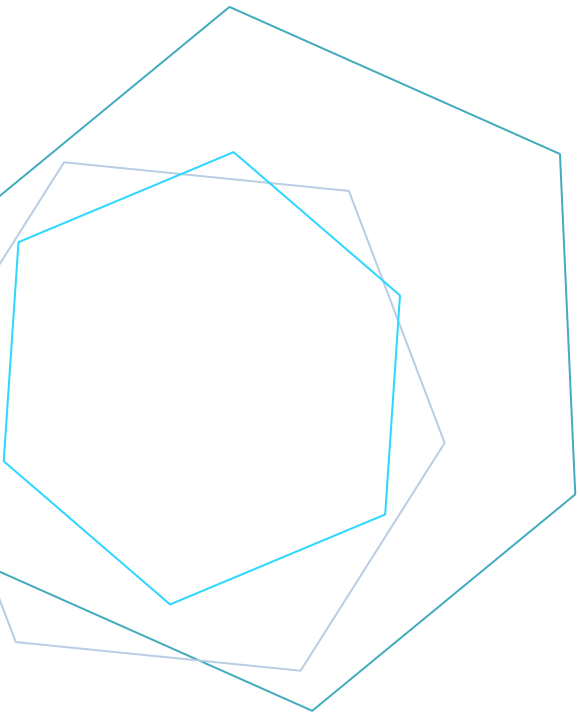
[2] Organizations listed in alphabetical order.

[3] We are also thankful for the support of the following organizations and people: Controladoria-Geral da União (Brazilian Office of the Comptroller General, CGU), Ministério da Ciência, Tecnologia e Inovação (Brazilian Ministry of Science and Technology. MCTI), Centro de Estudos sobre Tecnologias Web (Web Technologies Study Center, Ceweb.br), CodingRights, Data Privacy Brasil, InternetLab, Bruno Kunzler and Daniel Trielli.

Provided there is transparency and accountability, the implementation of AI technologies to improve public services may bring benefits to society. Without those premises, these technologies may have a huge negative impact on several rights, such as privacy, protection against discrimination, access to justice, freedom of speech, association and assembly, among others, as this document will point out. Moreover, the lack of legal standards allows for normative, regulatory, and ethical loopholes, due to the negative consequences of AI systems use without governance.

This work provides an overview of the AI tools currently used by the Brazilian Executive, Legislative and Judiciary branches in the federal level, for several purposes within the governmental scope. It also presents an analysis on their possible negative impacts on fundamental rights, as well as the main concerns pointed out by the civil society on the use of such technology.

This document seeks to summarize the analysis of negative impacts on fundamental rights carried out by representatives of the aforementioned civil society organizations. It concludes with governance recommendations for the public sector on employing AI systems.



CASE ANALYSIS

Initially, we mapped all AI algorithms and their current uses within policies and programs enacted by the Brazilian federal Executive branch. The methodology employed to obtain this information involved three steps: first, we conducted a survey with public administration bodies of the federal Executive branch; second, we sent Freedom of Information requests; finally, we searched official websites for updated information on AI use. These efforts resulted in a catalogue of AI tools used by the Brazilian federal public sector. (APPENDIX 01).

The survey was developed in partnership with Controladoria-Geral da União (Brazilian Office of the Comptroller General, CGU), Ministério da Ciência, Tecnologia e Inovação (Ministry of Science and Technology, MCTI) and Centro de Estudos sobre Tecnologias Web (Web Technologies Study Center, Ceweb.br) from Núcleo de Informação e Coordenação do Ponto BR (Brazilian Network Information Center, NIC.br). In September 2020, we sent the survey to the 319 bodies of the federal Executive branch listed in the Sistema Eletrônico de Informações ao Cidadão (Electronic System for Public Information Requests, e-SIC).

We also sent information requests to the federal Legislative and Judiciary branches, inquiring about the same information covered in the survey[4]. The requests were presented to 7

[4] All the requests, as well as the responses obtained, are available in Transparência Brasil's Achados e Pedidos project, which contains FOI requests, at: http://www.achadosepedidos.org.br/usuarios/tburg

bodies: Federal Senate, House of Representatives, Tribunal de Contas da União (Federal Court of Accounts, TCU), Conselho Nacional de Justiça (National Justice Council, CNJ), Superior Tribunal de Justiça (Superior Court of Justice, STJ) Supremo Tribunal Federal (Federal Supreme Court, STF) and Tribunal Superior do Trabalho (Superior Labour Court, TST).

Lastly, with Northwestern University support, we have developed an automated search algorithm that finds AI tools used by the public sector. The algorithm sifts through Google search results for keywords related to Artificial Intelligence, predictive models, and machine learning within websites whose domains are "gov.br", "leg.br", "jus.br" and "mp.br".

We then manually assessed 6,195 government site addresses (URLs) to determine if they were in fact related to AI use by the public sector. After this validation process, we extracted the text content from relevant websites and, by using the Scikit-Learn package in Python, we then developed a logistic regression model for predicting the terms most associated with matching results. This model, whose code is available in our Github repository[5], contributes to civil society efforts in identifying and monitoring new cases of AI technology use that happen to be mentioned in government websites.

Civil society organizations from various fields of expertise have contributed to building the framework for risk assessment on fundamental rights, to carry out a multisectoral analysis.

So far, we have mapped 44 AI tools used by federal government bodies. They have been grouped in two categories: the first classifies AI tools by whether they are used

[5] Link for the public repository: https://github.com/Transparencia-Brasil/algoritmos-brasil. The algorithm will go through further testing to ensure there is no bias. All the tests will be made available on the repository.

for decision making or not; the second takes into consideration whether the target group (final user) is internal or external to the public authority.

In the first category, a tool is considered to be used for decision making if it makes a decision autonomously or if it has been created to directly support a human decision. For instance, the Bem-te-vi tool from the Superior Labour Court categorizes processes and estimates proceedings in legal offices. In case it automates decisions that should be made by humans, it may affect proper access to justice and the fundamental rights of fair trial and due process.

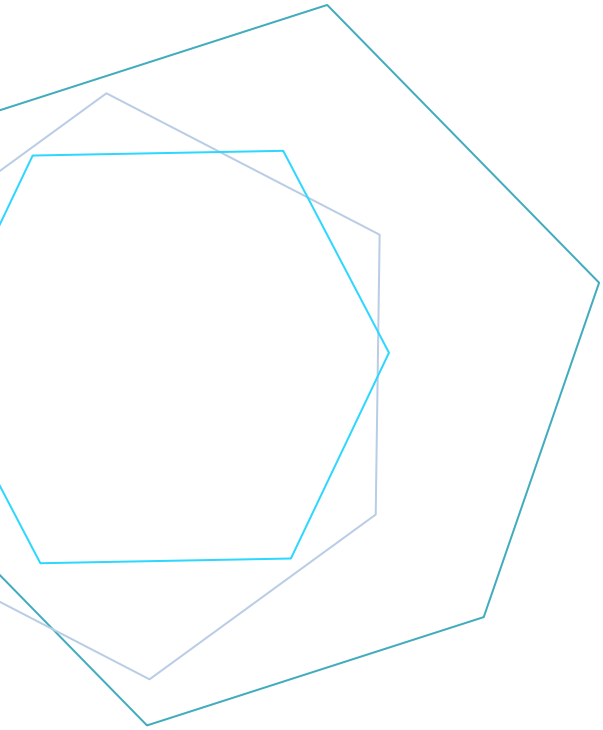
We define the algorithm as not influencing decision making if it is solely used to solve internal management issues unrelated to decision making processes, such as routine procedures automation. For example, the Federal Supreme Court Victor tool simplifies pattern recognition in textual data from federal cases: more specifically, it analyzes appeals and identifies the ones thematically linked to general repercussion rules. In this regard, the algorithm does not influence judicial decisions directly – it only optimizes internal procedures.

The second category classifies algorithms according to their target audience: either internal (governmental agents interacting with the tool) or external (citizens, companies and other entities impacted by the tool).

TABLE 1: CATEGORIZATION OF AI TOOLS ACCORDING TO DECISION-MAKING AND TARGET AUDIENCE CRITERIA.

TARGET AUDIENCE	WITH DECISION MAKING	WITHOUT DECISION MAKING	TOTAL
Internal	20 tools	16 tools	36
External	8 tools	0 tools	8
Total	28	16	44

Regarding the first category, 28 tools (64%) provide support in governmental decision making, whereas 16 tools (36%) are used to address internal demands that do not involve decision making. In relation to the target audience dimension, 36 tools (82%) are for internal use of civil servants working for the government body and 8 (18%) interact directly with external users (citizens/general audience).



FUNDAMENTAL RIGHTS RISK ASSESSMENT

When discussing and proposing governance recommendations for AI algorithms use, it is paramount to carry out risk assessments on real and possible threats to fundamental rights and to the civic space itself. It is also crucial to bear in mind that innovation and technological advances must always be aligned with accountability and transparency.

For this purpose, we suggest rating the impact of certain AI tools on fundamental rights in specific cases, based on their output or result – in other words, what they have been designed to deliver. It may also be about a possible error that may happen during its use.

The complete analysis is presented on the Fundamental Rights and Transparency Risk Assessment Framework: Artificial Intelligence Uses by the Public Sector[6], proposed by Transparência Brasil and including contributions from specialists and civil society organizations.

This classification is necessary since there are a myriad of objectives and goals within the governmental scope that might involve AI algorithms employment. Furthermore, these very tools tend to be quite diverse in their nature and complexity,

[6] Portuguese version available at: https://www.transparencia.org.br/downloads/publicacoes/Estrutura_Avaliacao_Risco.pdf

and thus might implicate quite different risks to fundamental rights.

Correctly identifying the risks these algorithms may pose to fundamental rights allows for mitigating them. Planning the design and implementation of the algorithm beforehand also contributes to preventing risks and making it easier to follow up on its operation and results. Moreover, it is possible to assess the relevance of having a specific regulatory framework covering all AI technological systems.

By crossing the typology data, as we can see on Table 1, we find three types of AI tools currently employed by Brazilian federal public authorities, according to Transparência Brasil mapping:

i) 20 tools to support decision-making aimed at public bodies
ii) 8 decision-making tools aimed at the external public, and
iii) 16 tools to improve internal procedures of public bodies, not involved in decision-making.

The tools employed by government bodies to support internal decision-making procedures make up most of the mapped cases. They also worry civil society organizations the most, since they aid government employees in taking certain measures which affect people's lives and may also impact, directly or indirectly, on their fundamental rights.

The main risk pointed out refers to database training and criteria used by the predictive models and automated ranking algorithms. If unadjusted, they may replicate existing prejudices and discrimination within the algorithm, producing biased results that may affect socially vulnerable groups.

One example is the PalasNET tool, employed by the Federal Police Department. According to them, the system registers confidential information from investigations and uses image and facial recognition techniques, as well as criminal profiling.

This tool was trained using a criminal database, from which little is known. If the database is not representative of the population, it may create discriminatory biases and increase the chances of obtaining false positives for certain groups, such as black or low-income population. Given the Brazilian criminal and justice systems are already highly discriminatory, this is no mere possibility, but a highly likely and worrisome scenario.

Another tool used in a different context but raising similar concerns is Weka, from Universidade Federal de Santa Maria (Santa Maria Federal University), which analyzes college dropout ratings. According to the department, the algorithm training variables come from academic databases.
This kind of tool might impact negatively on the right to education, since it might support measures that punish those ranked as having higher dropout probability (such as denying student financial assistance and social benefits).

Further negative impacts may arise due to including discriminatory factors within the analysis, such as considering financial constraints or maternity as factors that increase the likelihood of dropping out. This diminishes the chances of education for women, black people, and low-income people. Furthermore, it may also generate a self-fulfilling prophecy, since denying these benefits may itself increase dropout rates, thus 'confirming' the tool's predictions.

The second kind of tool supports decision-making and interacts directly with external users (such as citizens and the public). This is the case of chatbots, commonly developed to guide public service users, while reducing the dependency on human resources and increasing public service efficiency.

One of the risks these tools might pose relates to the level of accessibility of public services to those with reduced digital literacy – predominantly illiterate people, immigrants, or disabled people[7], – as well as those with unstable or unreliable internet access. Chatbots might be a gateway for accessing public provisions and goods, as well as for exercising several legal rights. Even if used properly, however, these systems might have negative impacts on citizens by not accounting for local traditions and culturally informed behaviors, preventing adequate access to public goods in certain communities.

For example, a chatbot for screening patients for Covid-19 symptoms, used by Hospital Universitário da Universidade Federal do Maranhão (Federal University Hospital in Maranhão), might generate negative impacts by perpetuating biases from preexisting screening issues – such as not recognizing a specific combination of symptoms as coronavirus infections.

It might also negatively affect minority groups with lower literacy rates, posing risks to public health.

Screening and estimating infection risk might turn out to be discriminatory depending on which criteria is used, favoring white people as well as middle and wealthy segments of the population.

[7] It is worth noting that the risk is not connected to one's characteristic or condition – as it is more evident in the case for disabled people: they might not be able to use a public service if the tool was developed without considering accessibility elements.

Furthermore, the very right to health might be jeopardized, since these tools work with a predetermined sequence of automated commands – based both on personal and non-personal data – which reach conclusions that could impact significantly on the life of any given citizen. In turn, citizens cannot promptly access the reasoning behind these tools, which makes it very difficult for them to understand why and how exactly it reached certain conclusions.

Finally, the third case refers to AI tools that work in simplifying internal procedures of the public sector. In other words, they are supposed to solve certain issues for government employees' workflow and to increase the capabilities of public administration agencies in computing high volumes of requests and demands. For this reason, this kind of tool has a lower risk of violating rights. If well employed, they might bring significant improvements for administrative procedures, especially in terms of increasing efficiency and reducing public expenditure.

For instance, the Brazilian Health Regulatory Agency (ANVISA) is developing a tool to facilitate the analysis of contributions to public consultations by grouping textual data. Another example is a tool employed by the Federal Court of Accounts Ombudsman, which classifies citizens' requests to facilitate comparisons and responses.

Even though such tools seem to have a lower potential for discrimination, one must always take into consideration the need for transparency and the responsibility of the public administration to abide by a set of fundamental rules, such as the principles of morality, legality, impersonality, publicity, and efficiency, all included in Article 37 of the Federal Constitution.

PUBLIC SECURITY

Tools for public security have a leading place in discussions on AI regulations around the world[8]. Although this document does not consider some rights as more important than others, it makes sense to tackle this subject, especially given the current popularity of AI tools in public security and the systemic racism in Brazilian society. We present two specific examples of AI technology used by the Brazilian government for public security purposes. We describe their risks and the impacts that must be taken into consideration in their implementation.

1. Facial recognition

Facial recognition systems are increasingly employed for public security purposes. In cases in which these AI tools widely scan public places, they affect everyone that passes through surveilled locations and might jeopardize the legal principle of presumption of innocence.

In addition, when not vetted for biases and properly corrected, they might impact negatively on the principle of non-discrimination. One example is Desk, an AI tool used by the Federal Police Department. Desk employs facial recognition and image processing and classification techniques to i) help to

[8] Amnesty International, for instance, has launched a campaign to ban its use: https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/

identify faces, ii) group faces by age clusters, iii) identify objects in places and iv) calculate the probability of nude content in a picture.

Our analysis has shown that the use of facial recognition for surveillance purposes always impacts negatively on the presumption of innocence, even if it would ideally work without any biases. The use of such technology reverses certain principles of democracy and rule of law. Specifically, it contradicts an established principle of criminal law, which says violating fundamental rights to privacy and personal data protection in the course of an investigation can only happen on the grounds of reasonable suspicion and probable cause. Facial recognition establishes indiscriminate surveillance of all citizens by a simultaneously ubiquitous and distant government entity, never allowing for civilian oversight or accountability of any kind.

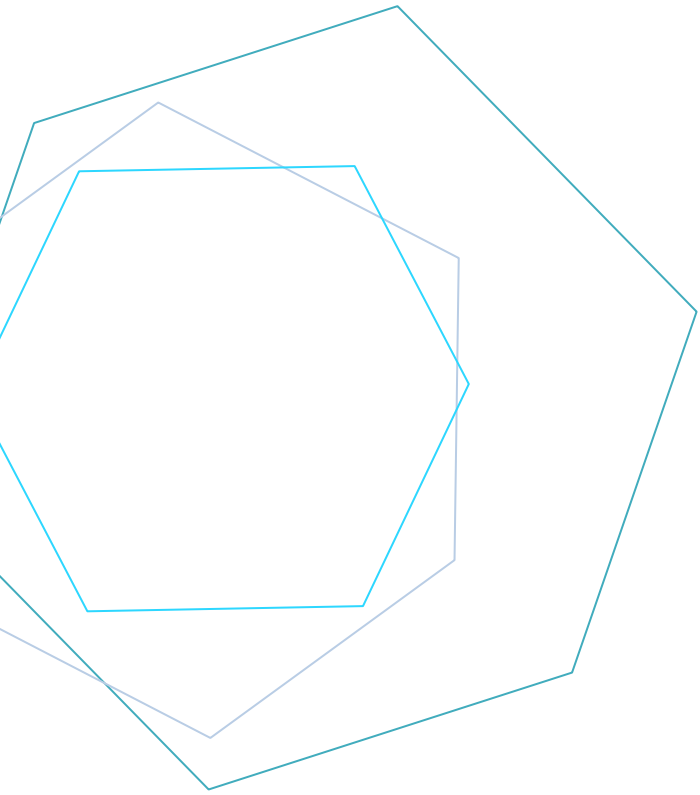
Moreover, the tool in question does not abide to the principles of transparency and non-discrimination. There are no indications of any guarantees or safeguards in its implementation and neither regarding the databases selected as training data for machine learning. Even if public security measures are exempt from the Brazilian General Data Protection Regulation (LGPD), they must still abide by its principles. The law states the need for further regulations on the matter. In fact, there have been international initiatives to suspend the use of these technologies until proper regulation is created[9]. As we will demonstrate in our recommendations, tools with such significant impact on fundamental rights should only be implemented after thorough impact assessment.

[9] The most famous case of suspending the use of facial recognition technologies happened in San Francisco, where legislators voted for the prohibition of its use towards their inhabitants. https://www.bbc.com/news/technology-48276660

2. Natural language processing

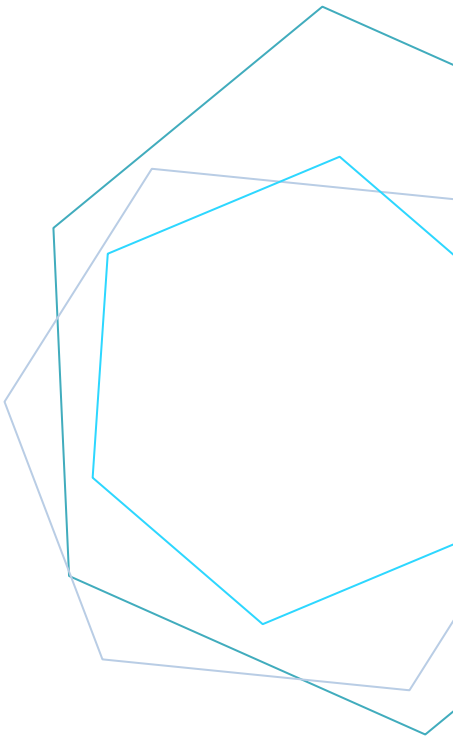
Natural language processing algorithms that estimate risks related to criminal matters may impact negatively on the rights of marginalized populations. For instance, the tool Localizador de Evidências Digitais (Digital Evidence Tracker), also employed by the Federal Police Department, uses natural language processing to estimate risks of criminal activity, such as fraud detection.

The algorithm may easily create biases against low-income people who either do not speak standard Portuguese or make heavy use of vernacular language, with abbreviations and slangs. This problem is not restricted to public security since biases against minorities might appear in any AI tool. But there are far worse consequences in the realm of public security since it might result in unjustified deprivation of liberty and severe threats to life.



GOVERNANCE RECOMMENDATIONS

Taking into consideration the aforementioned negative impacts on rights, we present four governance recommendations, summarizing the main concerns. In order to mitigate these risks, practical suggestions have been proposed for AI use and implementation by the public sector:



1. Representative databases

The goal is to prevent and eliminate/mitigate biases from the algorithms themselves and from the training data that may reinforce structural prejudices and violence (such as racism, sexism, LGBTQIphobia, among others), both in public provisions and investigations undertaken by public security agencies. The training database structuration is central to this analysis, not only since it might impact directly on rights to privacy and data protection, but also because it is a decisive step in which discriminatory biases tend to be produced.

According to contributions from InternetLab for the Artificial Intelligence National Strategy, carried out by the Ministry of Science, Technology and Innovation (MCTI), there are pragmatic ways to mitigate such discriminations – even though they might not be fully eradicated.

Biases are diverse and might originate from distinct sources. They may occur during data collection, cleaning, or treatment phases and/or during model-based testing[10].

During data collection, it is important to ensure the data selected for training the algorithms is representative of the groups and populations affected by them. It is also important to check for subtle discriminations embedded in data creation that might replicate discriminatory patterns further down the line. During the cleaning or treatment phases and model-based testing, pre-existing discriminatory beliefs from developers might contribute to perpetuating prejudices. According to InternetLab, *the main solution to avoid this bias is to have diversity in the project team*.

We also recommend that administrative and hiring processes for AI system development and implementation include legal devices and contractual terms obliging the use of databases which are representative of the affected population, besides requiring previous impact assessments. Lastly, the data chosen for model-based testing must be made available for external auditing.

2. Human supervision as safeguard for the revision of automated decisions

With respect to decision-making procedures for AI use by public authorities, the Law Act n. 13.709/2018, also known as Lei Geral de Proteção de Dados (General Data Protection Regulation, LGPD), states in Article 20 that the data subject has

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

the right to request the revision of decisions made *solely* based on automated personal data processing which affect their interests, including decisions regarding their personal, professional, consumption and credit profiles, or personality traits.

Article 20, §1° (first paragraph), states that the data controller must provide information on criteria and procedures used for automated decision-making in a clear and adequate manner, respecting trade and business secrecy as well as industrial confidentiality.

However, it is clear that the LGPD imposes the constraint that only decisions based solely on automated processing are eligible for revision. This constraint is a challenge to ensure safeguards on the obligation for human revision in automated decisions, because if interpreted in an extremely restrictive way, it may impair the fundamental right of individuals to control how their data are used and the impacts they may have in their lives.

The mapping shows results obtained through automated decision-making tools are employed to influence governmental decision making. However, one cannot assert that a certain decision has been taken exclusively by an AI once there is interaction between the systems and humans. It is hard to ascertain the level of support a given tool provides or even their level of automation.

The European General Data Protection Regulation (GDPR) ensures the objection to the decision made without human supervision. According to this regulation, data subjects must be informed about automated decision-making, as well as the

importance and consequences of the decision for them. GDPR also states that nobody can be subjected to decisions based solely on automated procedures that generate any effects over the legal rights of the data subject or affect them significantly. This ensures to data subjects the right to request human intervention and object to the decision.

We thus recommend that the Autoridade Nacional de Proteção de Dados (National Data Protection Authority, ANPD), responsible for protecting, implementing and supervising the LGPD, interpret Article 20 in a comprehensive way, since a restrictive interpretation of the term "solely" generates a scenario in which the right to review might never be effectively exercised.

This comprehensive interpretation has its basis on the very principles of the LGPD, as well as on the fact that the government is dealing with this data. Furthermore, we mention the results from international debates, such as the GDPR and the Santa Clara Principles on Transparency and Accountability in Content Moderation[11]. These normative instruments reinforce the need for human revision as a minimal criterion for effective protection of the subject data, in the context of automated decision-making processes.

It is important to highlight that during the LGPD approval procedures, the paragraph 3 of article 20 was vetoed. It stated the right of revision as a human right, in the sense that it must be exercised by a natural person. The human revision safeguard, however, is not guaranteed by the current Brazilian

[11] Santa Clara Principles on transparency and accountability in content moderation. https://santaclaraprinciples.org

law, which only mentions "revision on decision-making" – never making it clear which kind of revision it would be.

Thus, the law should include wide protections to guarantee the right to have algorithmic decisions reviewed by humans. This could be done through a new LGPD regulation or through the creation of a new, specific norm on the subject.

3. Effective personal data protection

Employing AI technologies requires processing large amounts of data for model training, boosting the construction and/or availability of large-scale personal databases, which might jeopardize citizens' right to privacy.

In this context, LGPD provides the following guidelines:

(i) Article 6 establishes that activities related to personal data handling must comply with good faith and the following principles: (...)

X – *liability and accountability for: indication that efficient measures have been taken by the agent, and capability of ensuring personal data protection regulation has been followed, including the efficacy of these measures.*

(ii) Article 23 establishes personal data handling by the government must fulfill its public purpose. Its item I states that "*one must report hypotheses in which (...) personal data are handled and provide clear and updated information on the legal provision, purpose, procedures and actions for the implementation of these activities, in easily accessible*

formats, preferably in websites". Furthermore, compliance with LGPD Article 6 is essential, and so is a non-negotiable respect for the principles of purpose, necessity, transparency, security, and non-discrimination.

This means data collection should only serve the purpose of ensuring public services provision or their improvement. Their purpose must be appropriate, well determined and based on the protection of fundamental rights.

People, as data subjects, need to be clearly informed of the purpose of personal data collection. It is important to understand how these government bodies store and use their data on their AI tools. Otherwise, this type of service is not compliant with principles established by LGPD.

ANPD must supervise personal data handling by the public administration. They also must supervise if an AI is employed in compliance with the purposes informed to data subjects and with lawful requirements of data privacy and protection – which will impose limits to algorithmic data processing.

4. Transparency and systems explainability

Accountability and transparency are necessary guarantees to enable the exercise of civilian oversight of Artificial Intelligence tools employed by public authorities, minimizing risks of fundamental rights violation.

It is paramount that government agencies and bodies promote transparency on their use of algorithms and automated

decision-making, if only for assessing the very efficiency of such tools.

Even if there are no threats to fundamental rights, one must always be able to assess how these technologies are being used and if there are actual advantages in it. They must also assess if the tool poses any harm, such as by highlighting social differences and imposing more oppression to already marginalized groups.

What do transparency and accountability mean for public algorithms? Above all, the need to ensure access to public information related to the algorithms employed by the public sector.

One possible way to ensure this is by using open-source AI algorithms. Codes must be followed by the description of the algorithm operation, the database used for training and, if possible, the main database in an anonymized version[12]. Besides, transparency mechanisms must be in place and citizens must be able to request a list of all algorithmic systems the public sector is employing, where and how, via Sistema Eletrônico de Informações ao Cidadão (Electronic System for Public Information Requests, e-SIC).

When there is restriction to access due confidentiality reasons (such as commercial confidentiality), there are two transparency and accountability devices that must be

[12] There are many ways to have data anonymization and this document is skeptical about the solutions that will be incorporated. For instance, public authorities may adopt differential privacy techniques to anonymize data, as well as create a synthetic database. For more about differential privacy and synthetic databases: 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, respectively.

considered in order to choose the best model. The government must inform at the very least the input and output variables of the model, as well as the algorithm type used (regression, neural network, decision tree, etc.). In other words, it is necessary that all phases of machine learning that result in a decision-making AI are easily traceable, and that variables responsible for the decision-making are made public.

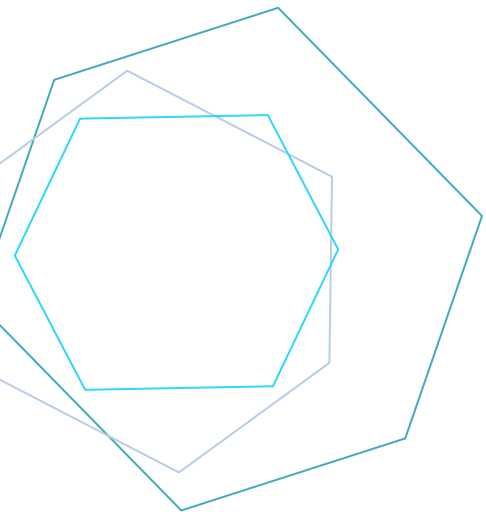
Another possibility for ensuring accountability is via auditing algorithms operations from time to time, carried out by experts external to both the hired AI company and the public body. This helps ascertaining if the algorithm is efficient enough for its purposes and if it causes any kind of negative impact on fundamental rights.

Thirdly, the government must ensure that algorithms are explained well enough: allowing citizens to understand how they work, how decisions have been made with their support, their purpose and reason, besides the data used in processing. It is important to note that effective transparency in these cases must also provide strong bases for oversight, supervision, and auditing of these systems by competent authorities, such as the Federal and State Prosecution Services and ANPD. Citizens might not be able to check AI results by themselves, even in cases in which there is enough transparency. Horizontal accountability is as important as the vertical one.

To ensure the ability to properly disclose and explain these AI tools, we suggest that the preparation and publishing of the Algorithm Impact Assessment before putting them to use becomes mandatory for all AI systems that might impact on fundamental rights, harm citizens and that use sensitive data, especially genetic ones. According to contributions to public

consultation for the elaboration of the Brazilian Artificial Intelligence Strategy, submitted by Instituto Brasileiro de Defesa do Consumidor (Brazilian Institute for Consumer Protection, IDEC)[13]: "Elaborating *and publishing preliminary reports also contributed for greater transparency of the AI system and for improving society's understanding of existing risks. This provides support for the concerned parties, the government and the general public to assess and determine if such risks are acceptable or not, as well as to make formal inquiries and complaints. People have the right to know how these AI systems impact their lives*."

Finally, the assessment carried out by civil society organizations pointed out that AI systems related to public security activities have higher risks to fundamental rights. The consequences of these risks are quite serious, involving deprivation of liberty of individuals or even life threats. Consequently, only strong public and civilian oversight measures might prevent these issues in AI tools developed for public security, especially the ones related to racist biases embedded in these tools. In case there is no specific legislation to regulate them, authorities must forbid the use of such technologies.



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

APPENDIX 01

Catalogue of AI tools used by Brazilian public sector

PUBLIC AGENCY
ANATEL
TOOL
Based on natural language processing, it identifies consumers' standard behavior, according to the users' complaint register to Anatel system and provides information for analysis
NEGATIVE IMPACT ON RIGHTS
Risk is related to the existence of other means for complaint. It is necessary that the tool is prepared to deal with different speeches, ways of expressing oneself and the issues.

PUBLIC AGENCY
ANTT
TOOL
Analysis of tweets
NEGATIVE IMPACT ON RIGHTS
There is the risk of the tool contributing for personalization and prediction of users' future behavior through tracks of their preferences left online. Such impact may restrain information access and relate to certain product consumption.

PUBLIC AGENCY
ANTT
TOOL
Prediction of daily average traffic in federal roads.
NEGATIVE IMPACT ON RIGHTS
It may impact on the right of freewill, as the criteria used by the model may lead to already existing social discrimination, or even generate facts. For example: the tool may suggest actions to direct people to places with more or less traffic, or more or less accidents. A place where there are more vulnerable people may have more accidents not necessarily because of the number of cars, but because there are more cars needing maintenance.

PUBLIC AGENCY
ANVISA
TOOL
Analysis of contributions to Public Consultations issued, by grouping text information
NEGATIVE IMPACT ON RIGHTS
The tool may impact on the rights to access to information and manifestation, depending on the way the algorithm has been trained and which elements the tool will actually take into consideration to group and prioritize the contributions. Both good and bad operations may impact on the way the agency listens to the civil society contributions in high impact issues such as permission for GMOs and pesticides, impacting on the rights for society to ponder on public policies.

PUBLIC AGENCY
Banco do Brasil
TOOL
Chatbot to answer regular questions on products and services
NEGATIVE IMPACT ON RIGHTS
In general, the fear about chatbot tools is about the access to services by illiterate people, migrants or those who do not have stable access to the Internet. Chatbot is the gateway for service and the exercise of several rights. There must be great care from the point of view of the speech used for it to be understandable and accessible, even if there is no favoritism from any side. Besides, if it is the main communication channel, it must be thought as inclusive, considering the possible connection elements (device, bandwidth,...)

PUBLIC AGENCY
Banco do Brasil
TOOL
Tool that recognizes if the face in a selfie provided to the bank app is the same as a picture from a personal document (drivers' license and ID) and identifies if the document is a drivers' license or an ID, once a digital account is created.
NEGATIVE IMPACT ON RIGHTS
There are risks related to the tool training and algorithm accuracy, having possible rejection consequences. The necessary cares are about data protection and information security.

PUBLIC AGENCY
Banco do Brasil
TOOL
Tool for automated screening for legal papers that identifies what kind of document it is and removes information such as emission date, county, state, expiration date.
NEGATIVE IMPACT ON RIGHTS
Even working well, the system may remove documents and impact negatively on local habits, denying access to public services.

PUBLIC AGENCY
Banco do Brasil
TOOL
Tool that predicts the probability of certain electronic transaction being fraudulent.
NEGATIVE IMPACT ON RIGHTS
One must consider possible biases that may impact negatively on low-income people or minorities. A more detailed analysis of the proxies for 'fraudulent operations' is necessary.

PUBLIC AGENCY
BNDES
TOOL
Tool for answering questions and guidance on emergency programs.
NEGATIVE IMPACT ON RIGHTS
Any bias in the language processing may negatively impact on minorities who don't write the language well, making it even more difficult to access services offered by BNDES. It is essential to establish a mandatory human revision of the outputs generated by the chatbot.

PUBLIC AGENCY
BNDES
TOOL
The tool identifies the best institution to fit the financing proposal, based on the analysis of the BNDES proposal profile.
NEGATIVE IMPACT ON RIGHTS
Even if this tool works well, it may cause exclusion, as the tendency is to direct the resources to those who have already had access to the Bank.

PUBLIC AGENCY
CADE – Conselho Administrativo de Defesa Econômica
TOOL
Tool to assess the probability of occurrence agreement between bidding companies.
NEGATIVE IMPACT ON RIGHTS
The model may impact on rights of consumers, investors and other companies. The background tendency must be to comply with greater agreements, which may generate misinterpretation. There may also be pooling of resources leading to possible issues concerning inclusion and access.

PUBLIC AGENCY
Caixa Econômica Federal
TOOL
Chatbot for guided conversations and bank transactions
NEGATIVE IMPACT ON RIGHTS
AI related to chatbots brings risks related to issues on speech recognition and accessibility. The language must be inclusive, driven to several contexts and groups. There may also be elements in architecture and the chatbot's flow that may restrain the exercise of other rights. The concern with chatbot tools in general is the access to services by illiterate people, migrants or those whose access to the Internet is unstable.

PUBLIC AGENCY
Caixa Econômica Federal
TOOL
Tool to detect fraud, suspicion of fraud, register blockage, authorization for debit and credit card transactions.
NEGATIVE IMPACT ON RIGHTS
The use of models to predict fraud may have a deep impact on people's lives. The care with specific elements that are analyzed is essential. Many times, models for fraud prediction take into consideration facts that not necessarily pose clear causal relationships. That way, they may harm people's expectations. The tool may impact on the right to access financial services and not authorize debit or credit card transactions based on discriminatory bias.

PUBLIC AGENCY
CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
TOOL
Tool to check the probability of i) two scientific works being the same one; ii) two non-identified people being the same person; iii) two organizations (public or private, international or national) being the same one; iv) two research projects being the same one or having the same investors
NEGATIVE IMPACT ON RIGHTS
Any bias or operating error in the scientific work analysis procedure may negatively impact scholars who depend on CAPES metrics to exercise their rights (example: progress in teaching career).

PUBLIC AGENCY
CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
TOOL
Tool to suggest recommendations on word search for journals.
NEGATIVE IMPACT ON RIGHTS
Any bias or operating error in the process of identifying a journal may negatively impact scholars who depend on CAPES metrics to exercise certain rights (example: progress in teaching career).

PUBLIC AGENCY
CGU - Controladoria Geral da União
TOOL
Tool to predict the probability of an agreement having problems with accountability.
NEGATIVE IMPACT ON RIGHTS
It is important to keep the need for a human to analyze the case, thus preventing the setting of any bias or operating problem the tool may have.

PUBLIC AGENCY
CGU - Controladoria Geral da União
TOOL
Tool to assess the probability of a certain case involving fraud or irregularity
NEGATIVE IMPACT ON RIGHTS
Occasional algorithm biases may favor or harm local administrators due to their political preferences. Depending on how it is trained, it may be useless to detect some frauds whereas it may rise false-positive in other cases, changing the focus of the investigation.

PUBLIC AGENCY
DPF – Departamento de Polícia Federal
TOOL
Register information from the confidential phase of investigations. The tool uses techniques for image recognition, facial recognition and criminal profiling.
NEGATIVE IMPACT ON RIGHTS
Tools for facial recognition may have several biases depending on the database used for its training. A possible bias is not recognizing black people or relating black people to criminal activities. These biases may impact negatively on the right to equal treatment. The tool may not suggest actions, but considering the techniques described it certainly rates people according to a more or less suspicious probability, which determines the course of the investigations. There is also a concern about the image database safeguard (facial biometrics is a sensitive personal data) and biases due to the way the algorithms are trained.

PUBLIC AGENCY
DPF – Departamento de Polícia Federal
TOOL
Tool for classifying images and estimating the likelihood that they contain nudity.
NEGATIVE IMPACT ON RIGHTS
Tools for facial recognition may lead to several biases according to the database used for their training. A possible bias may be about not recognizing black people or have the wrong perception on black people being related to criminal activities. The use of tools to detect nudity in an image may also have different impact according to the subject's gender. These biases may impact negatively on the right to equal treatment. It is also important that the national database containing child sexual abuse images may be diversified in order to mitigate the risk of these biases. Human review is certainly not done for all the images. In this case, a possible operating error would have a negative impact, by maybe not identifying violation of a children's fundamental rights.

PUBLIC AGENCY

DPF – Departamento de Polícia Federal

TOOL

Tool for facial recognition and image classification that i) helps identifying faces; ii) classifies faces according to age groups, iii) identify objects, iv) predicts the likelihood that the images contain nudity.

NEGATIVE IMPACT ON RIGHTS

If biases are not considered, it is possible that the tool generates misinterpretations and negative impacts on rights. Facial recognition tools tend to be less accurate in identifying and sorting black people. This mainly depends on the database used by the algorithm.

Considering the tool was fed with the national database of files containing child sexual abuse, one must analyze the different genre, ethnicity and race contained in this database in order to avoid possible biases.

Tools for natural language processing to predict some risk related to criminal matters may impact negatively on marginalized people. The algorithm may easily be biased against low-income people who do not know Portuguese well, or those who use informal language, abbreviations and slangs.

PUBLIC AGENCY

DPF – Departamento de Polícia Federal

TOOL

Tool that uses natural language processing to predict risks (including fraud detection). The tool helps to recognize entities (people's names, companies, addresses, figures, e-mails, telephone number, etc.).

NEGATIVE IMPACT ON RIGHTS

Systems for natural language recognition in public security, when not analyzed for possible bias detection, may impact negatively on the presumption of innocence principle.

Besides, even considering that public security activities are excluded from the Brazilian General Data Protection Regulation (LGPD) applicability, they need to comply with its principles. This tool does not follow the transparency and non-discrimination principles, as there is no evidence of the cautions that has been taken for its use.

It is an impact originated in the social environment, through possible biased databases, which is intensified by the usage of this tool. The fact that the system works better with texts written according to the formal language may lead to false positives or false negatives, harming people who have different educational levels. The absence of transparency on the criteria used to avoid possible biases may intensify the negative impact caused by the tool.

Natural language processing tools used to predict risks related to criminal actions may impact negatively on marginalized populations. The algorithm may easily be biased against low-income people who do not know Portuguese well or those who use informal language, abbreviations and slangs.

PUBLIC AGENCY
EBC – Empresa Brasil de Comunicação S.A.
TOOL
Tool to classify images in order to detect replication of images from TV Brasil in associate TV stations programs.
NEGATIVE IMPACT ON RIGHTS
No comments

PUBLIC AGENCY
EBSERH - HU-UFMA - Hospital Universitário da Universidade Federal do Maranhão
TOOL
Chatbot to screen patients who have Covid-19 symptoms. Tool to identify risk of infection caused by coronavirus.
NEGATIVE IMPACT ON RIGHTS
It's important to consider that biases and screening problems generated by the chatbot (for instance: not recognizing a specific combination of symptoms as a sign of coronavirus infection) may impact negatively on the right to health. Moreover, the right to health may be affected in case the pre-defined sequence for automated command, based on personal and non-personal data, come to conclusions that may have relevant impact on citizens' lives. The person may ask for additional information or immediate explanation on how the decision was taken. Depending on the criteria used, the screening and risk rate may be discriminatory, favoring white people and upper-class people, as previous experiences show.

PUBLIC AGENCY

Embrapa Agroenergia

FERRAMENTA

Tool developed in Tensor Flow to identify and classify vegetal species from pictures taken with drones.

NEGATIVE IMPACT ON RIGHTS

Although this tool is specific for vegetal species classification, eventual image capture by using drones over indigenous and quilombolas territory must involve previous, informed consultation, to the residents. The execution of projects, logs, and/or surveys in indigenous and quilombola territories without previous consultation to the residents violate regulations established by ILO 169.

This tool might have indirect impact on part of the population. It depends not only on the tool itself, but also on its distribution. There is also the possibility of error and lack of accuracy, but even when it works well, it might increase the gap between those who can access it and those who can't.

PUBLIC AGENCY

Embrapa Gado de Corte

TOOL

Predictive algorithm that informs the producer the best vegetable to be grown in their farms, based on agronomic information.

NEGATIVE IMPACT ON RIGHTS

Although this tool is specific for cattle pasture, it raises a question about what would be a recommended cultivation. This same logic may, in the future, be transposed to other cultures and have an impact on populations' food security. Those who suffer from 'productivity' pressure and are driven to pesticides usage, may in the future feel obliged to use predictive algorithms based on efficiency/productivity in the long run. This depends on how the algorithm has been set up and/or trained, considering what would be an adequate cultivation.

The bias here may tend for the logic of agribusiness and monoculture that doesn't consider traditional practices. The tool only takes into consideration agronomic measures matching the analyses from 5 specialists in obtaining results in the short run. Other types of knowledge must be considered if the system is meant for general usage.

In the specific case of cattle pasture, the question is if the tool also works to recommend the best cultivation for lands that have recently been deforested OR if the system will have any kind of restrain in offering information for lands in irregular situations, that is, that don't comply with committing to preserving the environment (where pasture is correct and where it is not).

This tool might have indirect impact on part of the population. It depends not only on the tool itself, but also on its distribution. There is also the possibility of error and lack of accuracy, but even when it works well, it might increase the gap between those who can access it and those who can't.

PUBLIC AGENCY
Embrapa Informática Agropecuária
TOOL
Tool to classify images that show plants diseases in some agricultural crops (tool to meet research request).
NEGATIVE IMPACT ON RIGHTS
This tool might have indirect impact on part of the population. It depends not only on the tool itself, but also on its distribution. There is also the possibility of error and lack of accuracy, but even when it works well, it might increase the gap between those who can access it and those who can't.

PUBLIC AGENCY
Embrapa Informática Agropecuária
FERRAMENTA
Tool to classify images for prediction of fruit-growing and fruit counting (fruit-farming accuracy) (under development).
NEGATIVE IMPACT ON RIGHTS
This tool might have indirect impact on part of the population. It depends not only on the tool itself, but also on its distribution. There is also the possibility of error and lack of accuracy, but even when it works well, it might increase the gap between those who can access it and those who can't.

PUBLIC AGENCY
Embrapa Trigo
FERRAMENTA
Tool that rates the probability of identifying diseases in wheat leaves (under development).
NEGATIVE IMPACT ON RIGHTS
This tool might have indirect impact on part of the population. It depends not only on the tool itself, but also on its distribution. There is also the possibility of error and lack of accuracy, but even when it works well, it might increase the gap between those who can access it and those who can't.

PUBLIC AGENCY
FURG – Fundação Universidade Federal do Rio Grande
TOOL
Chatbot that assists a certain audience to acknowledge the intentions when using the virtual learning environment.
NEGATIVE IMPACT ON RIGHTS
As it is aimed to a certain audience, it is important to understand if these people have been informed, in a transparent way, about the goal of the data collection and how the data has been used. That is, it is important to understand how the data has been stored, for the chatbot creation for this specific audience and if they have been informed about this storage. Otherwise, this service may be misusing the users' personal information. Besides, it would prevent third parties from having access to this information (once it selects a distinct profile) and influencing people the audience to make specific choices. The service is not verified by a human, which prevents the right to have enough and understandable information to understand the logic and criteria used for data treatment. If the chatbot output represents only a recommendation, not having human verification will not be a problem. In case the output represents a condition to access the virtual learning environment, though, the lack of human verification represents a violation to the right guaranteed by LGPD Article 20.

PUBLIC AGENCY
HFA – Hospital das Forças Armadas
TOOL
Tool to monitor patients' health conditions. Virtual screening to identify Covid-19 risk.
NEGATIVE IMPACT ON RIGHTS
Depending on the criteria used, the screening and risk classification may be discriminatory, favoring white and upper-class people, as previous experiences show. It's essential that there is human verification in this case. Any biased or problematic screening may impact negatively on the right to health. Besides, according to Article 20 of LGPD, any automated decision based on the treatment of personal data grants the information holder the right to request human verification.

PUBLIC AGENCY
INSS – Instituto Nacional do Seguro Social
FERRAMENTA
Tool for predictive analysis for patterns on the issuing of social benefits, seeking evidence of irregularities. The analysis made by the solution includes identifying deviations from standard behaviors expected in the analysis for granting certain benefits. The tool only identifies evidence of fraud and the corresponding amount from the analyzed data.
NEGATIVE IMPACT ON RIGHTS
No comments

PUBLIC AGENCY
UFRN – Universidade Federal do Rio Grande do Norte
TOOL
Tool to optimize the assistance to users by the ombudsman and the assistance for Freedom of Information requests.
NEGATIVE IMPACT ON RIGHTS
The possible tool bias (more accurate in relation to students, less in relation to the public servants) may impact negatively on the right to access information.

PUBLIC AGENCY
UFSM – Universidade Federal de Santa Maria
FERRAMENTA
Tool to predict the possibility of college dropout.
NEGATIVE IMPACT ON RIGHTS
If the tool is used for assisting decisions that punish those rated as having a high probability of dropping out (as denial for college financial aid), it may impact negatively on the right to education. Socioeconomic aspects must be included in this model - students who are going through financial difficulty but have good academic performance may also drop out. If the tool recommends actions, people on this group of students may no longer have access to education or benefits. If, with time, decisions are taken based only on the algorithm output, according to Article 20 of LGPD, the personal data holder (in this case, the student), will have the right to request human verification.

PUBLIC AGENCY

Tribunal Superior do Trabalho

TOOL

Tool for categorizing procedures and making predictions about the processing of documents in the Justices' Cabinets.

NEGATIVE IMPACT ON RIGHTS

If it automates decisions that should have been made by humans, the tool may hinder access to justice and to the right to due process. It may also compromise the legal duty of motivation for court decision. Even if, at first, the tool is used only to deliver a recommendation, it is necessary to analyze closely and with transparency, the way the algorithm is designed and the way it is used: the decision process cannot be dealt as a simple operation.

The suggestions made by the tool may influence the Judges' decisions. Consequently, if it includes previous decisions, it supports the non-development of jurisprudence, even if it works well. If it doesn't work well, and it suggests inappropriate solutions to the case, this can lower the quality of judgement. It may, thus, impact on access to justice, which works as an umbrella for labor rights.

It is important that the algorithm output remains as a suggestion. Any bias or problem in its operation may impact negatively on the right to a reasoned court decision or on the access to justice.

PUBLIC AGENCY

Supremo Tribunal Federal (STF)

TOOL

Tool to categorize legal procedures under general repercussion, which aims at simplifying the pattern recognition in legal texts submitted to the Federal Supreme Court.

NEGATIVE IMPACT ON RIGHTS

The algorithm may produce imbalance and inequality: someone who knows how it works will have more chances to have their appeal acknowledged. By applying reverse engineering, one can use keywords related to what the model understands as 'more likely to be acknowledged', thus creating manipulation. It is extremely important to grant broad transparency to the tool.

Categorizing legal procedures under general repercussion is a relevant aspect to judicial efficiency, especially when considering the role of STF within the Judiciary. Any bias or problem in the algorithm operation may impact negatively on the right to efficient and prompt judgement.

.

PUBLIC AGENCY

Superior Tribunal de Justiça (STJ)

TOOL

Tool that runs automated scan for each appeal presented to STJ and previous decisions of the suits, recommends ruling and legal precedent, recommends action (the final decision will always come from the STJ Justices).

NEGATIVE IMPACT ON RIGHTS

Considering the algorithm learns from previous rulings, it may replicate social prejudice included in Courts/decision makers. Although it doesn't decide, only "improves efficiency and speed for judicial assessment", its use may cause a false impression that the jurisprudence studies it delivers are neutral.

The same as previous cases mentioned applies: there is little transparency about the tool: will society have data on its usage? Do society know how often, and at which proportion the Justice rules based on the system's suggestion? How are these databases trained?

It is important that the algorithm output remains as a suggestion. Any bias or problem in its operation may impact negatively on the right to a reasoned court decision or on the access to justice.

PUBLIC AGENCY
Tribunal de Contas da União (TCU)
TOOL
Tool for text classification of PDF documents related to public funds losses, presented through the TCU's Tomadas de Contas Especiais (Special Audit of Accounts) management system (e-TCE).
NEGATIVE IMPACT ON RIGHTS
As previously mentioned, there may be a reinforcement of inequality between people who know the tool better and those who don't. In this case, it is important that the algorithm output remains as a suggestion. A possible bias or problem in its operation may lead TCU to not monitoring sensitive documents.

PUBLIC AGENCY
Tribunal de Contas da União (TCU)
TOOL
Tool to help with ruling accuracy for TCU, without investigated material errors. Its output includes notifications on inaccurate material (example: invalid social security number).
NEGATIVE IMPACT ON RIGHTS
A possible bias or problem in this algorithm's operation may compromise justice celerity, impacting negatively on the right to a prompt and efficient judgement.

PUBLIC AGENCY
Tribunal de Contas da União (TCU)
TOOL
Chatbot. Platform to facilitate the access to TCU's public solutions
NEGATIVE IMPACT ON RIGHTS
A possible bias or problem in this algorithm's operation may impact negatively on the right to access information.

PUBLIC AGENCY
Tribunal de Contas da União (TCU)
TOOL
Assisted instructions for legal opinions on rulings by Tribunal de Contas da União (TCU)
NEGATIVE IMPACT ON RIGHTS
As it is trained using previous decisions, the tool may replicate issues or decisions that negatively affect rights, especially for minorities who are usually disregarded or not appreciated by the court system decisions. A possible bias in this case may negatively impact on the right to access to justice and to the due process. It is important that any procedural phase takes into consideration all the information relevant to the case, even if it is only the preparation of a legal opinion.

PUBLIC AGENCY
Tribunal de Contas da União (TCU)
TOOL
Extraction of Justices deliberations on TCU rulings
NEGATIVE IMPACT ON RIGHTS
The possible impact on any rights depends on the purpose of the classification. If it is only to generate a suggestion, it doesn't seem to be a significant problem.

PUBLIC AGENCY
Tribunal de Contas da União (TCU)
TOOL
Classification of TCU's texts
NEGATIVE IMPACT ON RIGHTS
Tools to classify demands may impact on the right to justice. It is important that its outputs are considered only as suggestions and that there is human verification.

PUBLIC AGENCY
Tribunal de Contas da União (TCU)
TOOL
Tool to cross databases and produce graphic structures to analyze great volumes of data. The tool extracts the relationship between private individuals and entities within lawsuits and shows them for further analysis.
NEGATIVE IMPACT ON RIGHTS
Failure in reviewing digitalized documents may create arbitrary data selection and exclude certain lawsuits from the analysis that use the tool as source.

TRANSPARÊNCIA
BRASIL

transparencia.org.br