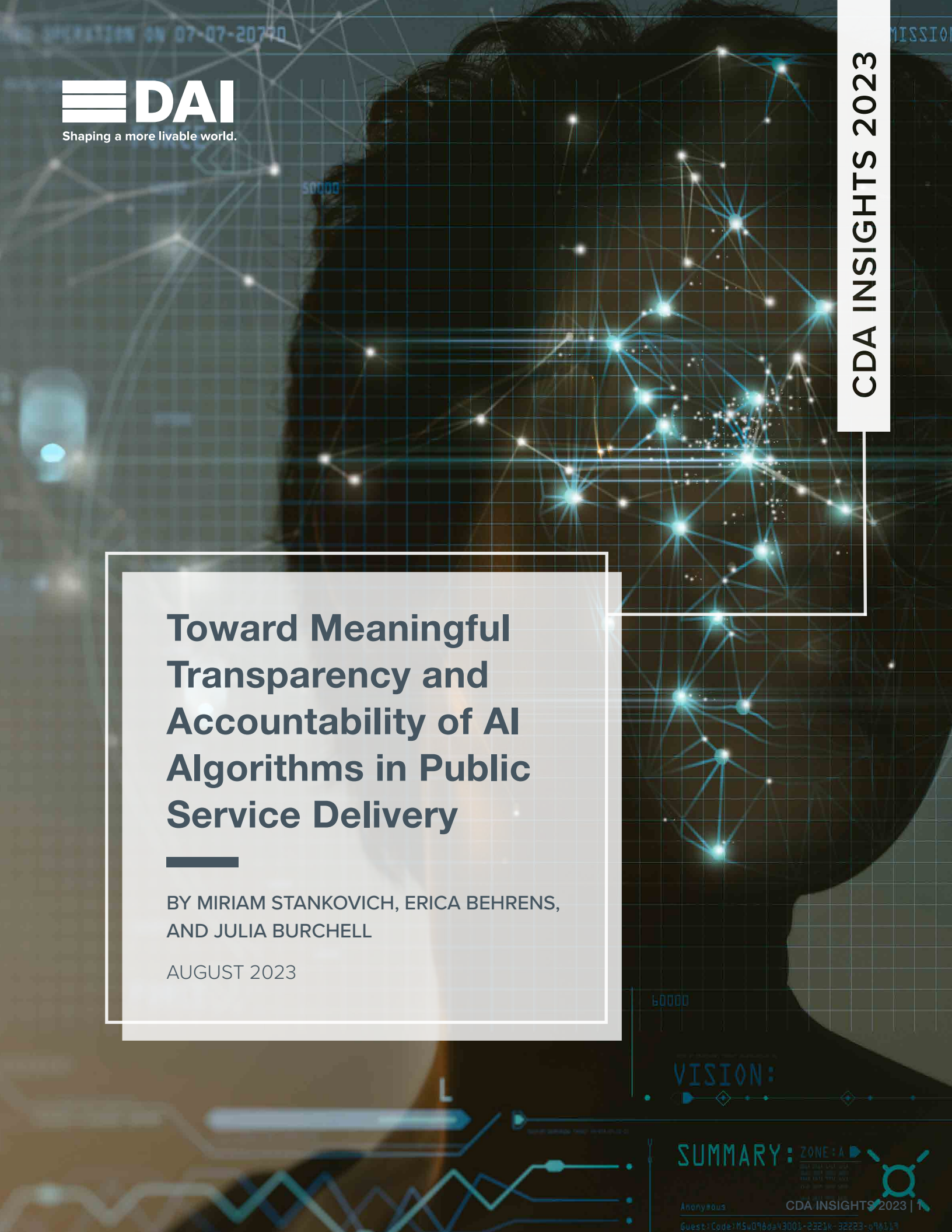


Toward Meaningful Transparency and Accountability of AI Algorithms in Public Service Delivery

BY MIRIAM STANKOVICH, ERICA BEHRENS, AND JULIA BURCHELL

AUGUST 2023

CENTER FOR
DIGITAL ACCELERATION

DAI's Center for Digital Acceleration helps our clients integrate digital tools and approaches across their portfolio, especially in emerging markets. We do this by engaging end users, building digital products, and understanding the broader ecosystems that drive the success of technology-based initiatives. Our clients include bilateral and multilateral donors, private sector companies, foundations, and others seeking to drive positive social change across sectors including health, governance, agriculture, education, and economic growth.

Contact Miriam Stankovich at miriam_stankovich@dai.com with any questions.

Design: Jennifer Geib {communication design}, www.jennifergeib.com

CONTENTS

Executive summary

Just a generation ago, the discourse surrounding "artificial intelligence" (AI) was mainly limited to academic disciplines such as computer science, philosophy, engineering, and science fiction. Today, AI is widely understood to be approaching the long-term goal of replicating human-like intelligence and is being incorporated in rapidly advancing technologies such as convolutional neural networks that will touch many aspects of modern society. These technologies are already being utilized in healthcare, national security, social media, agriculture, and various other fields.

The use of AI to harness data for innovation is increasingly important in driving economic growth and improving human well-being. In many parts of the world, AI is revolutionizing how people engage with technology. When ethically implemented, AI has the potential to tackle pressing global issues and offer substantial benefits to developing nations. Governments worldwide increasingly turn to algorithms to automate or support decision making in public service delivery—to assist in urban planning, prioritize social-care cases, make decisions about welfare entitlements, detect unemployment fraud, or surveil people in criminal justice and law enforcement settings.

However, AI systems in public service delivery can also cause harm, violate human rights (by reinforcing discrimination and undermining the privacy of digital personal data), and frequently lack transparency and accountability. AI systems are often deployed and work with sensitive personal data; these systems can make decisions that affect people's lives and might endanger fundamental human rights, such as the rights to privacy, freedom of expression, and access to information. Although AI systems are designed to be helpful, they may unintentionally amplify existing biases in society because they learn from the data on which they are trained, which may in turn reflect underlying social and cultural biases.

Further, AI tools are often regarded as "black boxes" due to their sheer complexity, making it difficult for human beings—developers included—to comprehend their reasoning. This opacity makes it challenging to question or probe AI functionality. **Since AI algorithms are involved in decision-making processes that affect lives, the consequences will be dire if this technology goes awry.** AI solutions must work for—not against—humanity as our processes become more automated.

This paper provides recommendations for those who seek to incorporate AI into public service delivery, drawing on interviews with experts in eight countries in Africa, Latin America, and the Caribbean that have witnessed increased deployment of automated provision of public services: Brazil, Chile, Colombia, Egypt, Ghana, Kenya, Mexico, and Rwanda. The recommendations emerging from this research are as follows:

- A **human rights-based approach** is essential to build and govern trustworthy AI systems in public service delivery.
- **Simplicity, context, and trust** are key to achieving algorithmic transparency in public service delivery.
- Addressing the **implementation gap** is fundamental to achieving algorithmic transparency in public service delivery.
- Any approach to addressing algorithmic transparency in public service delivery must be **tailored to local cultural, economic, and development contexts**.
- True accountability and transparency in algorithm usage by the public sector require a **multistakeholder approach and public-private partnerships**—fostering digital literacy, digital access, and digital rights awareness initiatives across the public, private, and civil society sectors.

Introduction

The use of AI to harness data for innovation is increasingly important in driving economic growth and improving human well-being. In many parts of the world, AI is revolutionizing how humans engage with technology. When ethically implemented, AI has the potential to tackle significant worldwide issues and provide substantial benefits to developing nations. From our social media feeds to the societal benefits we receive, AI is everywhere and affects everyone. AI may assist us in various ways: it can execute difficult, risky, or tedious tasks on our behalf; assist us in saving lives and coping with natural disasters; entertain us; and make our daily lives more enjoyable. AI helps doctors make decisions about our health and helps judges and lawyers sift through cases, speeding up the judicial process. With its cognitive, learning, and reasoning abilities, AI has the ability to boost industrial productivity, generate additional value for various industries, and anticipate and manage potential obstacles.

Governments worldwide increasingly turn to algorithms to automate or support decision making in public service delivery. Algorithms assist in urban planning, prioritize social care cases, make decisions about welfare entitlements, detect unemployment fraud, or surveil people in criminal justice and law enforcement settings. The use of these algorithms is often seen to improve efficiency and lower the costs of public services. AI tools such as chatbots, virtual assistants, and predictive analytics are on the rise in public administration, law enforcement, and the judiciary. This trend is accelerating because AI can process large amounts of data and recognize patterns between data sets, resulting in increased efficiency in administrative processes and improved delivery of public services.

The use of algorithms is often seen to improve efficiency and lower the costs of public services by automating or supporting decision making in public service delivery. However, AI systems in public service delivery can also cause harm, violate human rights (by reinforcing discrimination and undermining the privacy of digital personal data), and frequently lack transparency and accountability in their implementation.

Using AI in public service delivery: the experience of Singapore and Latvia

Latvia's UNA is a virtual assistant created by the Latvian Register of Enterprises and a private company specializing in AI solutions. Its primary objective is to enable Register clients to conduct legal subject and legal fact registration remotely. UNA replaces the need for in-person visits or calls to the call center and offers the convenience of 24/7 accessibility through clear, simple, and professional communication.

Singapore's Ask Jamie, first developed in 2014, is designed to assist citizens with queries related to specific domains and is active on 70 government agency websites. By enabling citizens to search for information online, Ask Jamie helps to reduce the workload of government staff who can then focus on more complex queries and issues. Ask Jamie uses a special AI technique, Natural Language Processing (NLP), to understand the user's query and provide an appropriate answer. In cases where the question has multiple permutations, Ask Jamie can ask follow-up questions to refine the answer to match the user's query.[1]

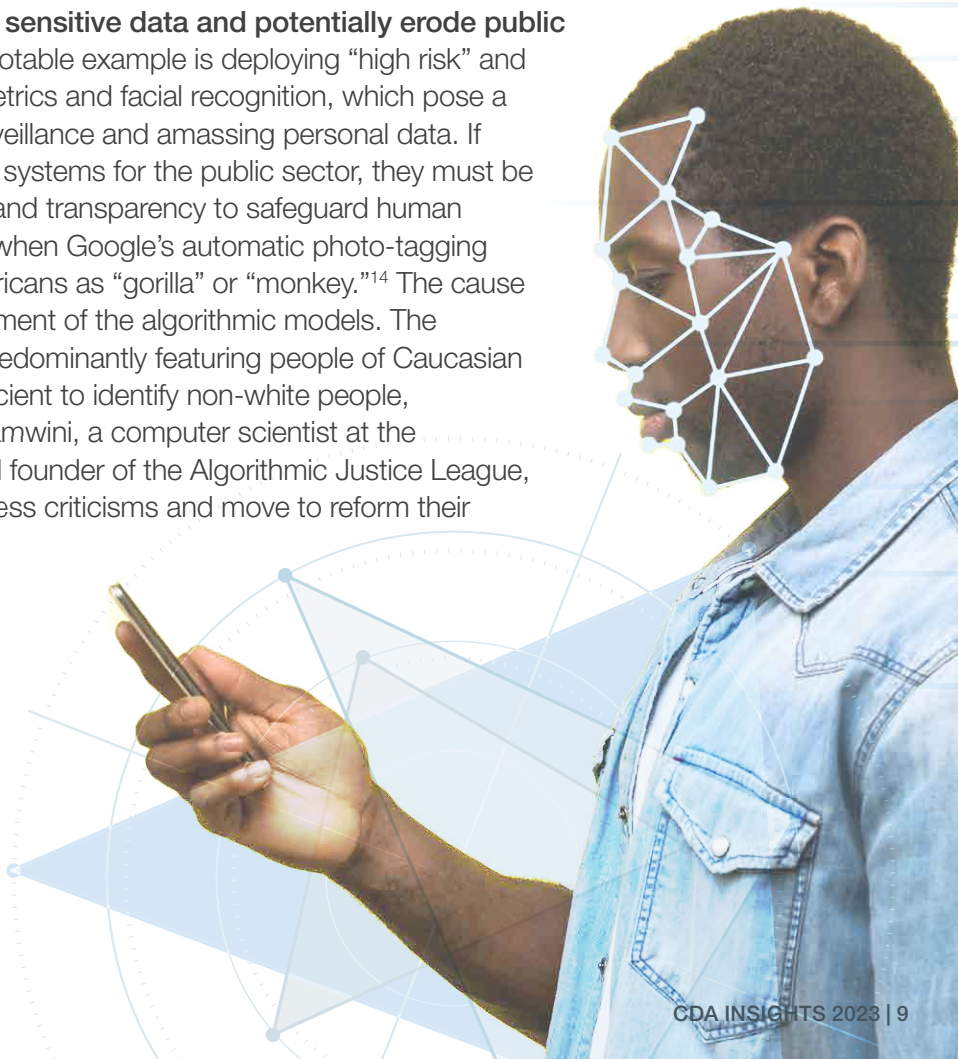
Public authorities are already using smart virtual assistants to enhance interaction with citizens. Some examples of this are Latvia's UNA[2] and Singapore's Ask Jamie.[3] The Indian Parliament uses AI to improve data processing and streamline legislative work.[4] Law enforcement agencies are also implementing facial recognition technology to aid in criminal identification. Additionally, judges and courts may explore the benefits of AI in hopes of achieving greater consistency in decision making[5] and promoting fairness in the justice system.[6] For instance, the Brazilian Supreme Court uses the VICTOR AI system, developed with the University of Brasilia. The AI technology analyzes the enormous volume of appeals brought to the Court. It automates the examination process by identifying cases with *repercussão geral* (general repercussion), a requirement to process an appeal before the Supreme Court.[7] In 2017, the Prosecutor's Office in the Autonomous City of Buenos Aires, Argentina, began developing Prometea. The AI tool has enabled the Prosecutor's Office to improve the efficiency of its processes significantly: for example, reducing tender processing time from 90 minutes to one minute.[8]

However, AI systems in public service delivery can also cause harm, violate human rights (by reinforcing discrimination and undermining the privacy of digital personal data), and frequently lack transparency and accountability in their implementation. AI systems are often deployed and work with sensitive personal data; these systems can make decisions that affect people's lives and might endanger fundamental human rights, such as the right to privacy, freedom of expression, and access to information. One recent example is a Colombian judge who used ChatGPT to assist in writing his rulings.[9] This and other instances of using AI technology without a proper understanding of its limitations have sparked debates and responses from regulators, and civil society watchdogs around the globe.

AI technologies work with vast amounts of data and can have cross-over and multiplicative effects that affect human rights issues and the rule of law. There is mounting evidence that AI systems are far from being neutral technology.[10] Instead, they can reflect their creators' (un)conscious preferences, priorities, and prejudices. Despite software developers' efforts to reduce their biases, bias can still be present in the data used to train an algorithm. Additionally, even well-designed algorithms must rely on imperfect and unpredictable information, making them prone to errors in new situations. Unfortunately, many developing countries have not yet started to educate the public about the use of algorithms in public service delivery or convene discussions on the topic.[11]

Since AI algorithms are involved in decision-making processes that affect lives, the consequences will be dire if this technology goes awry. AI solutions must work for—not against—humanity as processes become more automated. Consider the infamous instance when hundreds of eligible Dutch families were wrongfully implicated in fraud by an algorithm and made to repay social assistance.[12] The SyRI system risk indicator—used in the Netherlands to assess fraud risk—gathers and compares data from different government sources. A Dutch court recently concluded that the legislation establishing SyRI provided insufficient protection against intrusion in private life. The SyRI system lacked transparency, and its targeting of disadvantaged neighborhoods could amount to socioeconomic or migrant status discrimination. This illustrates the potential consequences of using biased algorithms in making decisions that impact our lives. In the field of policing, studies have shown that the use of predictive algorithms trained on past crime data replicates and amplifies existing systemic biases.[13]

Often, in the case of public sector use of AI systems, the technology is provided by a private entity. Public-private partnerships in the deployment of AI systems for public service delivery may give companies access to sensitive data and potentially erode public sector ownership of vital resources. A notable example is deploying "high risk" and "controversial" technologies such as biometrics and facial recognition, which pose a risk to human rights by enabling mass surveillance and amassing personal data. If private entities are contracted to deploy AI systems for the public sector, they must be subjected to the highest levels of scrutiny and transparency to safeguard human rights. To illustrate this, controversy arose when Google's automatic photo-tagging software identified pictures of African Americans as "gorilla" or "monkey."[14] The cause of these errors is likely to lie in the development of the algorithmic models. The models, trained with datasets of photos predominantly featuring people of Caucasian origin, had not been trained with data sufficient to identify non-white people, particularly women. The work of Joy Buolamwini, a computer scientist at the Massachusetts Institute of Technology and founder of the Algorithmic Justice League, has prompted multiple companies to address criticisms and move to reform their models.[15]



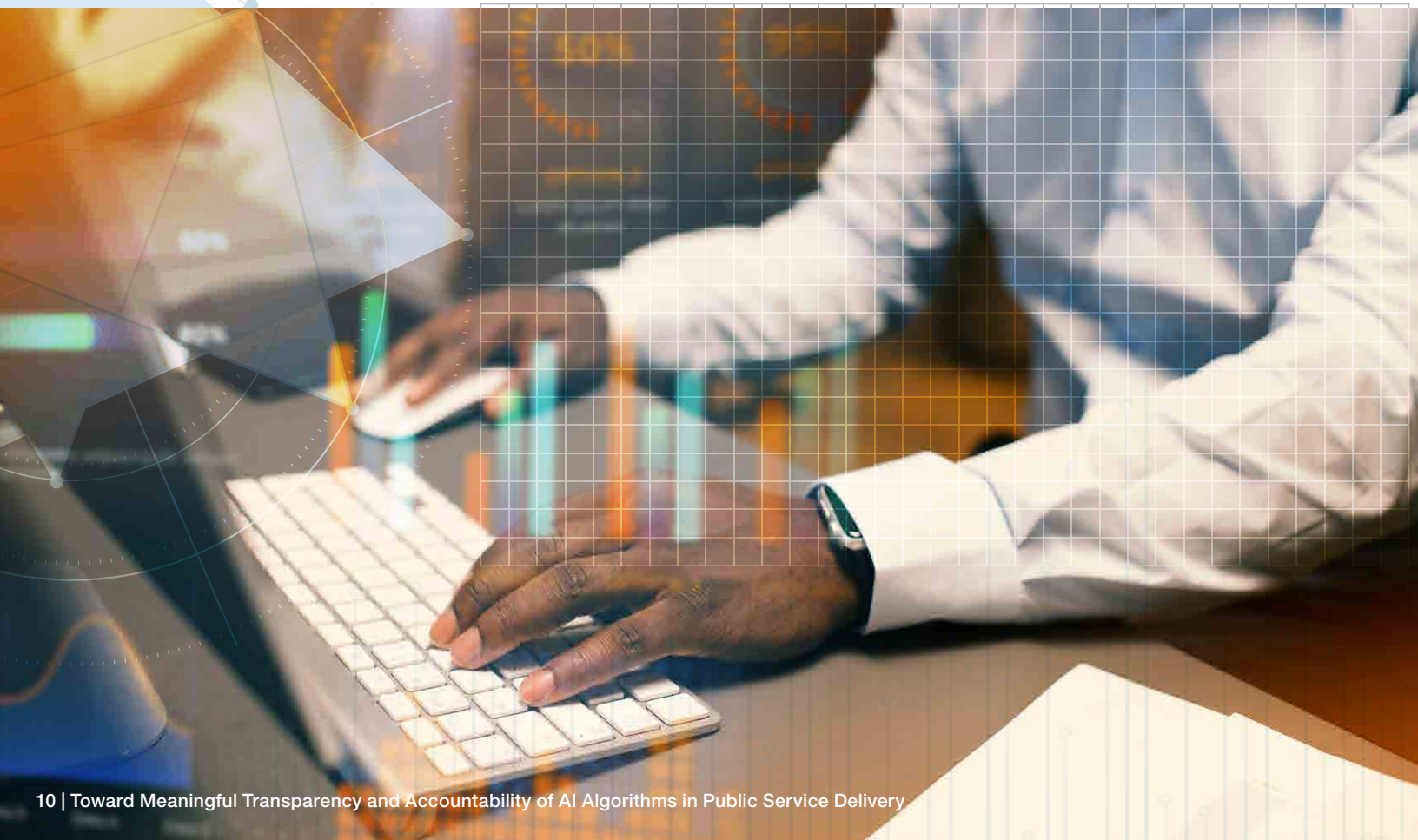
Many policy makers, regulators, and civil society organizations advocate for "algorithmic transparency" in public service delivery. Policies and regulations are implemented to address AI systems' transparency, explainability, and auditability issues; guard against discrimination; and enhance due process and personal data protection. Algorithmic transparency is about disclosing how algorithmic tools enable decision making by public policy makers and regulators by providing information in an open, understandable, easily accessible, and free format.[16]

While there have been efforts to evaluate algorithmic transparency within institutions or contexts in the "developed world," few systematic and cross-jurisdictional studies have been conducted to assess the implementation of these policies or their impact on human rights in developing countries. Existing literature in this nascent and fast-evolving space mainly focuses on developed economies. New research and the elaboration of an analytical framework are needed to generate informed insights applicable to developing countries.

This paper is the result of research conducted by DAI's Center for Digital Acceleration (CDA), offering insights that public policy makers and international development donors can use to:

- Identify when AI deployment in the provision of public services might affect human rights; and
- Include appropriate, proportionate, rights-ensuring "algorithmic transparency" elements in the delivery of public services.

The research is based on a literature review and key informant interviews in select developing countries in Africa, Latin America, and the Caribbean that have witnessed increased deployment of automated provision of public services: Brazil, Chile, Colombia, Egypt, Ghana, Kenya, Mexico, and Rwanda.



Framing the Problem

Algorithmic transparency and accountability in public service delivery

AI RISKS

AI technologies rely on vast amounts of data and can perpetuate biases related to race or gender despite efforts to minimize them. The data used to train algorithms can introduce bias, making it crucial to consider the risks and limitations of AI in various contexts.[17] This issue raises important policy, legal, regulatory, and ethical questions. For example, how can regulators and policy makers prevent algorithms from using biased or unfair data? How transparent should AI developers be about their decisions? Who has legal responsibility when algorithms cause harm? We see two foundational AI risks: algorithmic bias and AI black boxes.

Algorithmic bias

Although AI systems are designed to be helpful, they may unintentionally amplify existing biases in society because they learn from the data on which they are trained, which may reflect existing social and cultural biases. Additionally, even the most well-designed algorithm can make mistakes when faced with new or unexpected situations, a phenomenon often referred to as "artificial stupidity."[18] It is important to be aware of these limitations when using AI-powered technology.

Algorithmic bias is a crucial consideration when it comes to the use of AI and big data in development. AI is only as good as the data on which it is trained. This issue is particularly pronounced when AI applications are deployed in developing countries. Since many AI applications are designed outside the developing world, the available datasets are often sourced from individuals in developed countries. This can lead to a lack of accuracy in AI systems. AI systems may also embed algorithms that reflect the specific beliefs and biases of their creators. Such systems could contribute to discriminatory outcomes if applied in low-resource settings without input from the local population or access to local data.[19]

Algorithmic biases include stereotyping, prejudice, or favoritism towards things, people, or groups. These biases can impact data collection and interpretation, system design, and user engagement.20 Biases can arise in many ways in AI systems. Training data and AI models may be biased, or privileged groups may have advantages compared to other groups in AI decisions. Some of the most controversial biases in AI occur in facial recognition technology. A study conducted in Oakland, California in 2016 found that despite survey data showing an even distribution of drug use across racial groups, algorithmic predictions of police arrests were concentrated in predominantly African American communities, creating feedback loops that reinforced patterns of structural or systemic bias in police arrests.[1] Algorithms can also introduce racial biases when facial recognition algorithms are trained predominantly on data from Caucasian faces, significantly reducing their accuracy in recognizing other ethnicities.21 In 2016, ProPublica analyzed a commercial system designed to assist U.S. judges in making informed sentencing decisions by predicting the probability of criminals reoffending. The study revealed that the system exhibited bias against individuals of color.[22]

1 World Bank WDR 2021. The 2016 study conducted by the Human Rights Data Analysis Group using 2010 and 2011 data from the Oakland police department and other sources compared a mapping of drug use based on survey data from the victims of crime with another based on algorithmic analysis of police arrests. The study showed that biased source data could reinforce and potentially amplify racial bias in law enforcement practices. Data on arrests showed that African-American neighborhoods have on average 200 times more drug arrests than other areas in Oakland.

AI as a black box

AI tools are often regarded as "black boxes" due to their sheer complexity, making it difficult for human being—developers included—to comprehend their reasoning. This lack of transparency makes it challenging to question or probe AI functionality. Commercial AI developers usually keep their code private to maintain their proprietary intellectual property. This algorithmic opacity is alarming; creating accountable governance systems requires informed policy debate—something that is only possible with a clear understanding of how algorithmic systems operate.[23] In addition, the challenge of determining (human) liability for the actions of AI means that human rights harms can occur.

Algorithmic discrimination

Algorithmic bias and AI black boxes can lead to algorithmic discrimination. Both high- and low-income countries are susceptible to similar harms and threats from algorithmic decision making. However, the extent of this damage can vary significantly based on the presence of legal safeguards and accountability structures, particularly for marginalized communities.

Data invisibility and data justice as complementary to algorithmic transparency in AI deployment

Initiatives aimed at achieving data justice should complement any algorithmic transparency standards in government settings. A key challenge related to algorithmic transparency in the public sector is the exclusion of "data invisible groups." Those without digital access or who lack digital skills may not be included in assessments of populations and their needs.

"Data invisibility" results from the digital divide across many countries of the Global South and is likely to affect traditionally marginalized communities such as women, tribal communities, castes, religious and linguistic minorities, and migrant workers. In an increasingly digital world, data invisibility also means limited voice and reinforces restrictions on effective participation in social, economic, and political spheres. An overreliance on "automatic" data collection methods can exclude highly vulnerable groups and undermine trust in digital tools. Such exclusions may exacerbate biases that limit the effectiveness and validity of AI algorithms trained on easily accessible data—further reinforcing the need for greater transparency in data usage.

Governments and international development partners should promote inclusiveness and reduce digital inequalities by ensuring that data do not overrepresent those who are connected and give voice to data invisible populations.

Toward Solutions

Algorithmic Transparency and Accountability

ALGORITHMIC TRANSPARENCY

Demands for greater algorithmic transparency are growing in public and political debate, including requests that companies' algorithms should be reviewed by independent auditors, regulators, or the public before implementation. The initial step in implementing AI-based systems involves utilizing well-established algorithmic frameworks that ensure transparency and accountability. These frameworks should uphold data protection regimes and prioritize individuals' fundamental rights, such as privacy and the security of personal data. Many developing countries need greater regulatory and institutional capacity to address AI issues comprehensively. Transparency in automated decision-making requires the openness of datasets, communication of the analyzed data, and the mechanisms that the automated decision model employs.[24]

In an ideal scenario, individuals could comprehend and independently evaluate a system, even questioning its purpose and existence from the beginning. In this situation, people would not merely receive information on a system's functioning, usage, and associated organizational responsibilities but the societal role of the system and how it affects them.[28]

The call for transparency in algorithmic decision-making aims to address a persistent information gap between system users and operators. Algorithmic transparency measures how much information is made available to the user of the algorithmic system. Its scope encompasses an AI model's structure, its intended use, and how and when deployment decisions were made—and by whom. This also includes design decisions and training data used for the AI system. Without transparency, AI system users can only make educated guesses about these elements. (Of course, a more transparent algorithmic system is not guaranteed to be fair, secure, robust, or privacy-preserving).[29]

Algorithmic transparency in practice

The U.K. Central Digital and Data Office and Centre for Data Ethics and Innovation (CDEI) published one of the first national algorithmic transparency guidelines at the end of 2021. The standard is straightforward, consisting of a template that public sector organizations are encouraged to complete for any algorithmic tool that either directly engages the public (such as a chatbot) or meets specific risk-based requirements. The collected data is available in a public register. Public sector organizations piloted the standard, which was updated to reflect their feedback and will be reviewed every six months in the future.[25]

France, the Netherlands, and New Zealand have also developed guidance to help public sector officials navigate the responsible use of algorithms. France's Etalab supports government agencies in implementing a legal framework for accountability and transparency of public sector algorithms.[26]

Algorithmic transparency in the EU has been used at the local level since October 2020; the cities of Amsterdam, Helsinki, and Nantes, France established beta versions of registers describing the algorithms employed in city administrations. To ensure that AI used by public services is human-centered, the registers indicate, among other things, how data is processed, what dangers are involved, and whether the technologies are subject to human monitoring.[27]

Public access to entire algorithms or underlying code is rare, however, as private companies regard their algorithms as proprietary assets they are unwilling to disclose—often enforcing nondisclosure agreements. Despite these barriers, regulators have started demanding algorithmic accountability. For example, the EU GDPR requires companies to be able to explain how algorithms use customers' personal data and make decisions. According to Article 22 of the GDPR, EU citizens have the right to request that decisions made about them, which are based on automated processing and rely on their personal data, are taken by humans rather than only by machines. Additionally, individuals can express their opinions and challenge the decision in question.[30]

However, algorithmic transparency is not enough—even when decision-making processes supported by automated systems are transparent, subjects may feel that an algorithm has generated unfair outcomes and feel entitled to legal recourse. Unlike the comprehensive legal mechanisms that ensure responsibility for decisions made by human officials in the government—such as due process, appeals procedures, and laws that promote transparency—algorithms in public service delivery do not have similarly robust check-ins, unless required by law.[31] Decision subjects are left with little ability to rectify unjust outcomes due to flaws or oversight errors in the algorithm. Moreover, more data does not equal a transparent system because often data are not structured and lack quality. Ensuring adequate transparency of automated systems is complicated due to frequent algorithm changes. For instance, Google changes its algorithm hundreds of times per year. [32] Additionally, the risk of manipulating algorithms increases if they are made public.

"All AI systems must be designed to facilitate end-to-end answerability and auditability. This requires both responsible humans-in-the-loop across the entire design and implementation chain as well as activity monitoring protocols that enable end-to-end oversight and review."[33]

– The Alan Turing Institute (2019)

ALGORITHMIC ACCOUNTABILITY

Algorithmic accountability refers to the ability of those who design, build, procure, or implement the algorithm to be held responsible for their actions and impact according to policies and laws concerning algorithm use. A governing system holding an actor responsible requires that the actor be able to explain and justify decisions regarding the algorithm, and face consequences should those actions be against the law.[34]

The allocation of accountability for algorithmic decision making is complicated because frequently, it is not clear who or what has the necessary degree of control over the actions and outcomes of an AI system. Accountability is related to issues of owning responsibility. However, humans can only be accountable if they have a degree of control over an AI system, such as the ability to cause harm—prevent or mitigate it.[35]

Any regulatory system governing the use of algorithms in public service delivery must provide accountability if algorithms threaten to violate well-established human rights. Such a system must specify how accountability is to be fixed in the event algorithm-assisted decisions violate due process and fair trial requirements.

The public sector needs specific regulatory responses to ensure greater transparency and accountability of automated data processing and algorithmic decision-making systems. AI governance stakeholders need to institute mechanisms for safeguarding algorithmic accountability through due process and the rule of law. Effective redress mechanisms for individuals whose rights were violated by automated decision-making systems are also essential.

To ensure the efficiency and safety of AI-driven applications, AI governance stakeholders must maintain a human "in the loop; AI should not completely replace humans, but rather work in conjunction with trained professionals who can validate AI decisions. The effectiveness of AI relies on the quality of data, human capital, and expertise of the interdisciplinary team responsible for its development.[36]

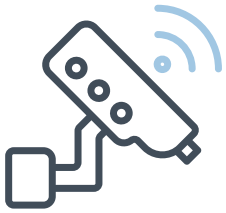
Algorithmic accountability in practice

Several local governments in the United States have implemented bans or temporary halts on utilizing algorithmic technologies, such as facial recognition technologies –(FRT), for law enforcement surveillance. The primary objective of these laws is to address concerns regarding privacy, but there are significant intersections with algorithmic accountability issues. These bans are typically established through legislation, but some laws have provided limited exceptions to the prohibition, such as third-party information obtained through FRT. For instance, a bill in San Francisco that prohibits using FRT only applies to use by municipal agencies and excludes usage by federal agencies, such as those in ports and airports. [37]

One of the pioneering policies on algorithmic accountability is the Canadian Directive on Automated Decision-Making (ADM), which includes an Algorithmic Impact Assessment (AIA). Any Federal public agency must conduct an AIA before producing any ADM system, whether at the project design stage or just before production. These AIAs must be updated when there are changes in functionality or system scope and made available to the public.[38]

Insights and Recommendations from our research

Key recommendations for international development partners, policy makers, and regulators are summarized below:



For AI to benefit the public good, its design and implementation must avoid harming fundamental human values.

A human rights-based approach is essential to building and governing trustworthy AI systems in public service delivery.

The lack of algorithmic transparency and its impact on human rights are at the forefront of discussions on AI. AI systems in public service delivery raise concerns about thoroughly assessing their short- and long-term effects, whose interests they serve, and if they are adequately sophisticated to deal with challenging social circumstances in developing countries. Due to a lack of knowledge and access to AI systems, producing convincing solutions to these important concerns has been difficult. Certain AI applications in the public sector might be incompatible with human rights—for instance, using biometric or facial recognition systems in public spaces might enable mass surveillance. According to the AI Now Regulating Biometrics report (2020), facial recognition technology is not an adequate identification replacement for fingerprints. Facial recognition technologies show poor performance results and high error rates for "black women, gender minorities, young and old people, members of the disabled community, and manual labourers."[39] Government deployment of AI might encroach on due process and equal protection rights if, for example, AI is used for DNA testing, criminal justice risk assessments, or watch list selection. As such, AI technologies might impact human rights, such as freedom of thought, conscience, and religion; freedom of expression and information; and freedom of assembly and association.[40]

For AI to benefit the public good, its design and implementation must avoid harming fundamental human values. International human rights law provides a robust framework for these values. Human rights law, as an international framework, is intended to establish global principles and mechanisms of accountability for the treatment of individuals. A human rights-based approach provides normative guidance to AI developers to uphold human dignity, regardless of jurisdiction. Implementing human rights can aid in developing technical and policy safeguards in AI deployment.

Human rights law is a globally acknowledged value system based on the rule of law. It is an established means for ensuring the protection of rights in general and in the digital environment, including the rights to equality and non-discrimination. Its nature as a universally binding, actionable set of standards lends itself especially well to technologies that transcend national boundaries, such as AI.

A human rights-based approach is essential to build trustworthy AI systems in public service delivery. To ensure this approach in public sector operations, developing countries' governments should have a readily accessible analytical framework to assist them in identifying when AI components might impact human rights and how algorithmic accountability could mitigate those risks. Where AI systems threaten fundamental rights, countries should protect and promote those rights and ensure that private sector actors conduct due diligence and human rights impact assessments (HRIAs) according to their responsibility. The outcome of HRIAs should lead to different safeguards assigned to the specific risks and impacts established in the process. [41]

Governments around the globe, such as the United States (Blueprint for an AI Bill of Rights),[42] have attempted to address AI accountability and transparency issues from a human rights perspective. The fundamental rights and algorithm impact assessment (FRAIA) tool, developed by the Dutch Ministry of Interior and Kingdom Relations, provides a valuable framework for conducting algorithmic impact assessments based on the human rights approach[43]

Most government institutions tend to use pre-made algorithmic products that they purchase or license from third-party vendors when deploying algorithmic systems. However, they often lack the necessary tools to assess vendors' proposals beyond comparing costs. To account for potential harms and address specific risks before implementation, frameworks are needed to evaluate how proposed systems function and perform across various categories of harm.

The CDA team has developed an analytical framework for algorithmic human rights impact assessment that international development donors and government officials can use throughout the lifecycle of AI tools to evaluate their impact on human rights in automated decision-making processes. A literature review and key informant interviews confirmed the need for this type of assessment framework. One interviewee from Colombia indicated that public policy officials and government representatives need more specific guidelines that go beyond ethical frameworks for AI and automated decision-making systems. Instead, public sector officials need to understand AI development and deployment in public service delivery through an international human rights lens. Many public sector officials and regulators struggle with understanding and unpacking broad AI principles such as transparency, fairness, and accountability. Officials often ask "What does this mean? Or, how do I then implement this in practice?" CDA's research indicated that contextualizing is very important. We cannot take certain models deployed in high-income settings and transpose them into developing settings. Countries in Africa, Latin America and the Caribbean, and Southeast Asia face different cultural and sociological challenges that are hard to measure.

The CDA team has developed an analytical framework for algorithmic human rights impact assessment that international development donors and government officials can use throughout the lifecycle of AI tools to evaluate their impact on human rights in automated decision-making processes.

The analytical framework for human rights impact assessment of algorithms offers direction for developing and using transparent and accountable automated development systems for public service delivery by providing a flexible, simple, organized, and quantifiable process to handle AI transparency and accountability risks across the AI lifecycle. The assessment's goal is to support international development partners and developing countries' public sector officials in managing transparency and accountability risks associated with creating, implementing, and using automated decision-making systems. Adopting the framework can help and guide these stakeholders to understand and decide what levels of risk are acceptable.

The guiding principles of the framework are the following:

- The transparency and accountability of AI systems can be achieved through public disclosure of information about the system in question, its processes, its direct and indirect effects on human rights, and what actions have been taken to identify and mitigate adverse consequences.
- In all cases, the information provided should enable a meaningful evaluation of the AI system.
- No AI system should be so complex that human assessment and inspection become impossible.
- If an AI system cannot meet adequate transparency and accountability standards, it should not be used in public service delivery.[44]

PHASE 1:

Mapping, Conceptualization, and Initial Analysis

Phase 1 will involve the government institution officials that intend to use the AI system, its developers, as well as representatives of the population/demographic cohort that will be affected by the implementation of the AI system.

PHASE 2:

Design, testing, and implementation

Phase 2 will rely on the results of Phase 1. If Phase 1 does not indicate any red flags, the process could proceed to Phase 2. Phase 2 will also be inclusive, and it will involve the government institution officials that are using the AI system, its developers, as well as representatives of the population/demographic cohort that will be affected by the implementation of the AI system.

PHASE 3:

Monitoring and evaluation

Phase 3 will involve the government institution officials that are using the AI system, as well as/or any other auditing organizations that specialize in algorithmic audits or algorithmic impact assessment.

TABLE 1. Analytical framework for human rights impact assessment of algorithms[2]

PHASE 1: Mapping, conceptualization, and initial analysis

What will the automated decision do?	• Who is the algorithm's intended audience, and who will be most impacted by the automated decision making? (e.g., children, women, minorities, etc.). • What is the nature of the algorithm used for automated decision making? (Is it a non-self-learning algorithm in which humans specify the regulations the computer must observe; or a self-learning algorithm, in which the machine finds patterns in the data?). • Do we have sufficient training data to generate accurate algorithmic predictions regarding the decision? • Are the data used for training sufficiently varied and trustworthy? • What is the algorithm's data lifecycle? • Which groups will be impacted the most by training data errors, and discriminatory treatment?
What are the objectives of the automated decision-making process?	• Why is the algorithm needed and what outcomes is it intended to enable?
What is the legal basis for automated decision making?	• If an algorithm is expected to affect human rights, what is the legal basis for its use?
What methods will be used to identify any possible biases?	• Can you provide information on the testing process for the AI algorithm, including the target audience for testing and how algorithmic bias will be measured and corrected, particularly concerning marginalized communities?
What are the incentives for automated decision making?	• What will be the main benefits from the algorithm's development? • What are the possible adverse outcomes, and how will we identify them? • How transparent will we make the algorithm's design process to internal partners and external clients? • What action will be taken if it is predicted that the development or deployment of the algorithm may result in undesirable outcomes?
How are additional stakeholders engaged?	• What is the algorithm's feedback loop for developers, internal partners, and customers? • Do civil society groups have a part in the algorithm's design? • Does academia have a part in the construction of the algorithm?
Has the design and implementation of the AI system taken diversity into consideration?	• Will the algorithm affect particular cultural groups and behave differently in cultural contexts? • Is the design team sufficiently diverse to capture cultural subtleties and foresee the algorithm's applicability in various cultural contexts? If not, what measures do we have in place to make these scenarios more prominent and comprehensible to designers? • Considering the algorithm's objective, are the training data sufficiently diverse? • Are there statutory guidelines that public sector organizations should check to ensure that the application of the algorithm is legal and ethical?

2 This framework has been adapted from the following sources: Brookings, Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms; Singapore AI Governance Framework (2020); Government of the Netherlands, Fundamental Rights, and Algorithms Impact Assessment (FRAIA).

PHASE 2: Design, testing, and implementation

Infringed fundamental rights	Are the design testing and implementation going to impact fundamental rights such as: • Right to access to court. It is important to acknowledge that AI systems have the capacity to embed bias, which can hinder access to the courts and justice for marginalized groups whose data may not be visible. • Fair trial and due process. Automated decision-making processes by AI systems can have a significant impact on individuals. However, these systems often do not provide individuals with the chance to participate or contest the outcome of the decision. Furthermore, many AI systems are not capable of presenting an explanation of their decisions in a way that is easily understood by humans. • Privacy and data protection. The use of AI/ML heavily relies on collecting and processing digital data by tracking individuals' online behavior, which can impact their right to privacy and data protection.[45] • Freedom of expression. Digital platforms use automated algorithms to manage third-party content during political and electoral campaigns. While their intentions may be good in identifying and removing 'extremist' content, there is a concern that these actions may not adhere to legal, legitimate, and proportional standards for permissible interference with freedom of expression. • Effective remedy. When AI systems are used in situations involving human rights, it can be challenging to ensure that individuals have the right to remedy. This is because many AI systems are opaque and people may not know how decisions affecting their rights were made, or whether the process was discriminatory. Additionally, judicial operators using AI systems may be unable to explain the automated decision-making process. These challenges are magnified when machine learning systems that recommend, make, or enforce decisions are used within the justice system, which is responsible for ensuring rights, including the right to an effective remedy. • Rights to protection against discrimination. There is a possibility of infringing on certain rights due to the significant risks of bias and discrimination that come with implementing machine learning algorithms. These biases may arise from algorithm developers, the model upon which the systems are built, the data sets used to train the models, or the introduction of biases when such systems are applied in real-world scenarios. • The right to explanation[46]. Understanding the outcomes produced by AI systems can be challenging for end-users as these systems combine and recombine attributes in seemingly arbitrary ways. To address this challenge, the right to explanation has been established, necessitating that end-users can interpret and comprehend the specific elements utilized in the ML model responsible for each outcome. Achieving explainability poses a technical challenge for AI developers.[47]
Specific legislation	• Is there a specific legislation that limits the design, testing and implementation of the algorithm?
Seriousness of interference	How seriously is a fundamental right affected by the algorithm? • Serious interference, thus compelling reasons required as justification (red) • Medium-serious interference, thus due diligence required (yellow) • Less serious interference, thus no special due diligence required (green) • A useful risk-based assessment framework is provided by the draft EU AI Act.

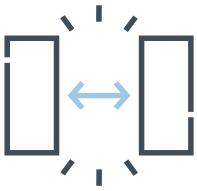
PHASE 3: Monitoring and evaluation

Level of human involvement It is essential to be able to measure the level of risk and impact of AI systems that might be deployed in public service delivery. It is crucial to determine the requirement for human oversight based on the use case, its sensitivity, the complexity and opacity of the algorithm, and the potential impact on human rights – whether this implies the human is "in the loop", "on the loop" (HOTL), or "in command" (HIC).[3]	• What role do humans play in decision making based on the algorithmic output? • Is there a human in the loop? • Is there an active and involved human oversight, with the human retaining full control and the AI only providing recommendations or input? For example, a judge may use AI to evaluate certain aspects of a case. However, the judge will make the final decision. In the case of human out of the loop, a criminal recidivism solution may automatically rank individuals based on pre-determined demographic and behavioral profiles • Does the AI model provide enough information for the human to make an informed decision (e.g., factors that are used in the decision, their value and weighting, correlations) • How is staff empowered to make decisions responsibly based on the algorithmic output? • Is there sufficient qualified staff in place to manage, review and adjust the algorithm, if needed, and will there be in future?
Internal processes safeguards	• How often and at which moments in time is the use of the algorithm evaluated? • Does the organization have the right staff in place to do so? • Regarding self-learning algorithms: have processes and systems been set up to monitor models (e.g., with respect to data drift, concept drift and accuracy)?
External processes safeguards	• Is there a mechanism for external auditing and supervision in place?

3 According to the European Commission High-Level Expert Group on AI: "HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal, and ethical impacts) and the ability to decide when and how to use the system in any particular situation." (European Commission 2021). See also Personal Data Protection Commission, Singapore (2020).

Simplicity, context, and trust are key to achieving algorithmic transparency in public service delivery

International development partners, policy makers, and regulators should have a balanced approach to accomplishing algorithmic transparency and accountability in public service delivery commensurate with the risk and workload for civil servants. Simplicity, trust, and context were the top factors echoed in interviews. If the algorithmic HRIA frameworks are not easy to understand, public sector officials have difficulties implementing them. This was the case, for example, with the Digital Republic Act in France, which requires transparency for certain public algorithms. Still, at first, agencies found it difficult to comply, partly because of a lack of resources and precise instructions.[48]



Realizing and addressing the implementation gap is instrumental in achieving algorithmic transparency in public service delivery.

As one Colombian interviewee said, "*Ethical frameworks and even laws won't do much good if they are little more than words on a page.*" Developing countries face greater constraints in implementing fair, accountable, and transparent algorithmic systems due to lower levels of development. Donors and international partners need to meet governments where they are, supporting holistic approaches and practical programming that supports deploying ethical and human-rights based systems for algorithm usage.

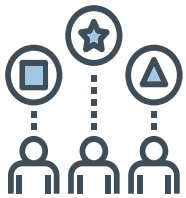
The appropriate building blocks must be in place before algorithmic deployment in public service delivery can be successful.

This holistic approach should encompass the following building blocks: (i) sound algorithmic governance (including accountability mechanisms), (ii) meaningful connectivity so that data invisible groups and remote populations are not left behind, and their data is fed into the AI systems that will be used for public service delivery, (iii) comprehensive data protection and sharing mechanisms, and (iv) public understanding of where and how data is being used and how legitimate inquiries about algorithmic outputs may be redressed. The appropriate building blocks must be in place before algorithmic deployment in public service delivery can be successful.

According to our research, different governments have different priorities regarding the building blocks outlined above. In Egypt, for instance, data quality and governance are a high priority, but transparency is a low priority. Without government initiative and little understanding or demand from the public, the lack of transparency opens the door for government misuse or abuse of data and algorithmic systems.

In Ghana, interviewees cited a lack of government strategy and policies for AI usage, as well as a lack of quality data that are clean and usable for analytical or algorithmic purposes. While the use of algorithms for public service delivery is not yet widespread in Ghana, the lack of data interoperability, poor transparency in data usage, and weak regulations and policies undermine the ability to implement useful and fair algorithms.

In the case of Colombia, the OECD notes that the country developed a monitoring dashboard of AI systems used by public entities.[49] While the dashboard is free and available to all citizens, one of our interviewees, a professor at the University of Rosario in Colombia, explained that there were several AI and algorithmic systems known to him through his research that did not appear on the dashboard, due to the complicated and burdensome nature of inputting a new entry. This discrepancy between policy and real-life execution is one example of this implementation gap.



The collection and use of data and rollout of algorithmic systems must account for a country or region's specific cultural and demographic context.

Approaches to addressing algorithmic transparency in public service delivery must be tailored to local cultural, economic, and developmental contexts.

Like implementation in developed countries, the collection and use of data and rollout of algorithmic systems must account for a country or region's specific cultural and demographic context.

This sentiment was perhaps most prominent among interviewees in Kenya, who separately raised concerns regarding inclusion and fairness concerning gender, tribal affiliation, marginalized communities and ethnic groups, and geographic location.

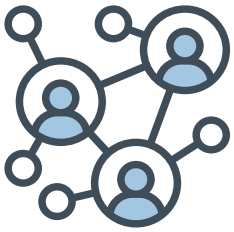
A thoughtfully designed, inclusive algorithm deployed in one country cannot be replicated in another with the same results. Additionally, while extensive knowledge of algorithmic use is not widespread, several interviewees cited high demand for data protection, rights, and regulations among Kenyans, demonstrating a foundational awareness of digital rights among the public.

In Ghana, with a population of 33 million and more than 50 languages spoken, language was cited as a major barrier to AI uptake and implementation in the country, both by government and non-government actors, and is an issue in the implementation of other digital tools as well. Though English is the country's official language and *Akan* is widely spoken, various regional languages are still used in schools and the

delivery of public services, presenting significant barriers to inclusion for many. For example, the Bank of Ghana is developing a chatbot incorporating the *Twi* language. English language-based platforms can have a pre-disposed bias towards specific demographic cohorts (such as geographic location or level of education). Public and private sector service providers look to ensure inclusive service offerings by integrating natural language processing and accommodating for first language.

The Mozilla Foundation has also been working to address inherent language bias in tech platforms. Its Common Voice platform is working to diversify contributors to voice-enabled platforms. For example, Mozilla is working to increase participant use among women and rural populations to build out more nuanced AI functions for use cases in *Kiswahili*.[50]

One interviewee from Kenya explained that she still struggles with using Google Assistant, Alexa, and Siri because the platforms do not understand her Kenyan accent. While these platforms have been around for nearly a decade, they still have not demonstrated machine learning to accommodate different dialects, accents, and cultural nuances.



True accountability and transparency in algorithm usage by the public sector necessitates a multistakeholder approach, public-private partnerships and fostering digital literacy, digital access, and digital rights awareness initiatives among the public, private, and civil society sectors.

When major components of the larger digital and data ecosystem are lacking—such as digital and data literacy, infrastructure, and collection policies and standards— an algorithm can produce incomplete or harmful, biased outputs. In addition to a sound governance framework concerning the use of algorithms, governments must prioritize support for the health of this larger ecosystem. If not, top-down efforts for transparency, accountability, explainability, and fairness can only go so far. Given the still-developing nature of many digital and data ecosystems, countries face challenges in operationalizing fair, transparent, explainable, and accountable algorithms.

We are beginning to acknowledge that algorithms have both positive and negative effects on our daily interactions. The challenge now is to work together to improve the design of algorithms, comprehend their impact, address any negative consequences, and guarantee that their expected beneficial results are distributed fairly. Governments cannot design policies solely for specific public sectors like healthcare, social services, or

The challenge is to work together to improve the design of algorithms, comprehend their impact, address any negative consequences, and guarantee that their expected beneficial results are distributed fairly.

education. Success in implementing robust algorithmic transparency and accountability frameworks in the public sector will require collaboration among multiple stakeholders to ensure that AI solutions can be integrated into developing contexts. To deploy AI and big data for development, policy interventions ensuring affordable devices and data services, bandwidth, and energy are critical. It is essential to create local content and applications in local languages, enhance digital literacy skills among different stakeholders, and develop AI, data engineering, coding, economic, and creative capacity. By using the proposed analytical framework for human rights impact assessment of algorithms, potential harms can be identified, and actors can work together to determine the necessary policy, ethical, and technical measures needed to fix these issues.

Our research and key informant interviews show that countries have different approaches to multistakeholder governance of algorithmic transparency and accountability in public service delivery. For instance, according to our Mexican interviews, the current government has relatively little interest in algorithms, but universities and startups are still pushing forward initiatives in algorithmic transparency. An example of a public policy prototype for transparent and explainable AI systems is Open Loop Mexico, which is based in Mexico. Open Loop Mexico aims to provide public policy recommendations for more trustworthy AI through collaboration with Meta, the IDB, the Mexican Data and Transparency Regulators (INAI), and 10 Mexican AI companies.[51] In Brazil, the Centre for the Fourth Industrial Revolution Brazil (C4IR Brazil), a think tank and spin-off of the World Economic Forum (WEF), implemented the WEF AI Procurement in a Box toolkit in partnership with *São Paulo Metrô* and the *Hospital das Clínicas*. According to a WEF report[52] it is crucial for developing countries' governments to conduct thorough self-assessments of IT infrastructure, data practices, and institutional maturity before adopting AI technologies. The report also recommends policy measures like algorithmic impact evaluations and certifications. C4IR Brazil and the WEF advise governments to be aware of the interaction between humans and AI to avoid algorithmic and human-based bias. The hospital pilot focused on the ethical use of AI in healthcare, given its public scrutiny and the institution's technical and data access. The *Metrô* pilot, on the other hand, faced the challenge of implementing an innovation procurement procedure for the first time and used an algorithm impact assessment tool to identify and mitigate potential risks in deploying a predictive maintenance AI tool.[53]

GobLab Chile: an example of public-private partnership in algorithmic governance

The University Adolpho Ibanez, through the GobLab, and the Chilean Council for Transparency are working together on projects including data science, transparency, data protection, and ethical problems. This collaboration is important in light of the expanding data management and automation capabilities, such as artificial intelligence, algorithms, and others, that assist decision making in the public sector in Chile.

The project is working on the development of a registry of automated and semi-automated decision systems deployed in the Chilean public sector using data voluntarily submitted by public organizations under the Council for Transparency's oversight. This will make it possible to learn whether and how this technology is being used in various government agencies. The second stage will analyze a collection of systems and assess their transparency using a framework of international standards that has been modified to comply with Chilean law. As a result of these efforts, it would be possible to assess how well a normative suggestion for ensuring the algorithmic openness of public institutions was developed.

With funding from the Inter-American Development Bank's Innovation Lab, GobLab later drafted and proposed a regulation that the government is on track to adopt following initial testing of the regulation with various public bodies. The regulation will make Chile the first nation in Latin America to adopt standards on algorithmic transparency.[54]

Digital literacy matters in algorithmic transparency and accountability. Even when algorithms deployed in public service delivery are transparent, individuals still need the ability to redress and understand them—and use available resources to remedy the situation. The potential risks of AI are visible in Uruguay, where the Ministry of the Interior had to discontinue the use of a predictive policing algorithm after a trial because the public officials using the system could not fully grasp its impact on human rights. They opted instead for retrospective statistical analysis methods. This incident highlights the importance of data literacy among public sector officials when implementing AI solutions for public service delivery. It is crucial to assess AI alongside other less advanced—but more suitable methods—of data analysis and use.[55] Governments and international donors should prioritize the digital literacy, rights, and access efforts of civil society and marginalized groups. Governments unable to assure accountability and transparency of an algorithm should reconsider if it is the most appropriate solution—or incorporate nondigital means of explanation and redress.

As reflected in the proposed assessment, it is important for civil society organizations to be included in the first two phases of implementation. This step will enable them to advocate for the inclusion of marginalized groups in the algorithm design process and ensure that important values are accurately reflected in the code. An interviewee from Fundación Karisma, a Colombian civil society research and human rights advocacy organization, cited lack of trust in government systems among both government actors and the public.[56] Carrying out research in digital autonomy and dignity for years, the group has recently investigated the use of government algorithms for public services, finding a lack of transparency and examples of bias in their deployment. With a team of lawyers and extensive research, the group works to hold government agencies legally accountable for algorithmic misuse.

Annex I:

Key legal and policy instruments related to AI, ethics, and human rights

NATIONAL LAWS AND POLICY INSTRUMENTS

Non-exhaustive list:

Draft EU AI Act (2021)

The EU Parliament Resolution with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics, and related technologies (2020)

Proposed UK AI rulebook

Proposed US Algorithmic Accountability Act of 2022

China Regulation of Recommendation Algorithms (2022)

EU Digital Services Act Package

Canadian Directive on Automated Decision-Making (ADM)[57]

UK Algorithmic Transparency Standard

REGIONAL AND INTERNATIONAL LAWS AND POLICY INSTRUMENTS

Non-exhaustive list:

The European Convention on Human Rights

Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data

Council of Europe, Recommendation CM/Rec(2020) of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.

Guidelines of the Committee of Ministers of the Council of Europe on online dispute resolution mechanisms in civil and administrative court proceedings (2021)

European High-Level Expert Group on AI: Ethics Guidelines for Trustworthy AI (2019)

UNESCO Recommendation on AI Ethics (2021)

The Council of Europe's Ethical Charter on the Use of Artificial Intelligence in Judicial Systems

Universal Guidelines for AI (2018)

OECD AI Principles (2019)

G20 AI Guidelines (2019)

Africa CPHR–Resolution 473

High-Level Expert Group on AI, Ethics guidelines for trustworthy AI

SOFT LAW INSTRUMENTS AND INNOVATIVE/AGILE APPROACHES TO AI GOVERNANCE

FAT/ML Principles for Accountable Algorithms and a Social Impact Statement for Algorithms

IEEE (Institute of Electrical and Electronics Engineers) Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design*

There are examples of emerging standards by industry associations such as the IEEE on algorithms, transparency, privacy, bias and more broadly on ethical system design and the Internet Engineering Task Force (IETF):

- IEEE P7000: Model Process for Addressing Ethical Concerns During System Design
- IEEE P7001: Transparency of Autonomous Systems
- IEEE P7002: Data Privacy Process
- IEEE P7003: Algorithmic Bias Considerations
- IETF Research into Human Rights Protocol Considerations draft

The Montreal Declaration for a Responsible Development of Artificial Intelligence

The Asilomar AI Principles, developed by the Future of Life Institute

C Minds algorithmic transparency policy sandbox in Mexico in partnership with Meta

Algorithm registers (Amsterdam, Helsinki)[58]

The Japanese Society for Artificial Intelligence Ethical Guidelines (2017)[59]

Algorithmic impact assessments

- Canada, Algorithmic Impact Assessment Tool
- Singapore's AI Verify

ENDNOTES

1 https://www.tech.gov.sg/products-and-services/ask-jamie/; https://oecd-opsi.org/innovations/una-the-first-virtual-assistant-of-public-administration-in-latvia/

2 https://www.ur.gov.lv/en/about-us/una/

3 https://www.tech.gov.sg/products-and-services/ask-jamie/

4 https://www.financialexpress.com/india-news/ai-push-parliament-assemblies-to-use-artificial-intelligence-and-machine-learning-under-e-vidhan-project/1426380/

5 https://www.malaymail.com/s/1711625/cj-artificial-intelligence-for-sentencing-virtual-hearings-holograms-in-tom

6 https://www.wired.com/story/algorithms-shouldve-made-courts-more-fair-what-went-wrong/

7 https://www.conjur.com.br/2021-jul-09/salomao-braga-judiciario-agenda-2030-onu; https://portal.fgv.br/en/news/artificial-intelligence-judiciary-and-its-role-implementing-un-agenda-2030; https://sifocc.org/app/uploads/2020/06/Victor-Beauty-or-the-Beast.pdf.

8 https://unescochair.cs.uns.edu.ar/en/2020/06/prometea-transforming-the-administration-of-justice-with-artificial-intelligence-tools/; https://ceridap.eu/prometea-experience-using-ai-to-optimize-public-institutions/?lng=en; https://www.ibanet.org/article/14AF564F-080C-4CA2-8DDB-7FA909E5C1F4

9 https://www.law.com/nationallawjournal/2023/03/13/chatgpt-helped-write-a-court-ruling-in-colombia-heres-what-judges-say-about-its-use-in-decision-making.

10 https://www.americanbar.org/groups/judicial/publications/judges_journal/2021/winter/artificial-intelligence-benefits-and-unknown-risks/

11 https://www.transparency.org/en/blog/algorithms-artificial-intelligence-public-administration-transparency-accountability

12 https://algorithmwatch.org/en/syri-netherlands-algorithm/

13 https://towardsdatascience.com/fighting-back-on-algorithmic-opacity-30a0c13f0224; https://iapp.org/news/a/digital-welfare-fraud-detection-and-the-dutch-syri-judgment/; https://pace.coe.int/en/files/28715/html

14 https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement

15 https://www.ajl.org/

16 https://www.gov.uk/government/collections/algorithmic-transparency-reports

17 https://www.americanbar.org/groups/judicial/publications/judges_journal/2021/winter/artificial-intelligence-benefits-and-unknown-risks/

18 https://www.dai.com/uploads/ethical-ai.pdf

19 https://www.frontiersin.org/articles/10.3389/fdgth.2020.00006/full

20 https://developers.google.com/machine-learning/glossary/

21 Hill, Kashmir. 2020. “Wrongfully Accused by an Algorithm.” New York Times, August 3, 2020. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

22 J. Angwin, J. Larson, S. Mattu, L. Kirchner, Machine Bias, ProP, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

23 https://digitalregulation.org/3004297-2/; https://dai-global-digital.com/can-regulators-keep-up-with-artificial-intelligence.html

24 Ibid.

25 https://cdei.blog.gov.uk/2022/06/01/piloting-the-national-algorithmic-transparency-standard/

26 https://www.opengovpartnership.org/stories/open-algorithms-experiences-from-france-the-netherlands-and-new-zealand/

27 https://algoritmeregister.amsterdam.nl/en/ai-register/; https://ai.hel.fi/en/ai-register/; https://data.nantesmetropole.fr/pages/algorithmes_nantes_metropole/

28 Ada Lovelace Institute (2020) Meaningful transparency and (in)visible algorithms

29 https://advance.sagepub.com/articles/preprint/Transparency_in_Predictive_Algorithms_A_Judicial_Perspective/14699937

30 https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/my-rights/what-are-my-rights_e

31 https://law.yale.edu/sites/default/files/area/center/mfia/document/algorithmic_accountability_report.pdf

32 https://searchengineland.com/google-seo-news-google-algorithm-updates

33 Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529

34 https://www.opengovpartnership.org/wp-content/uploads/2021/08/executive-summary-algorithmic-accountability.pdf

35 European Parliament "A governance framework for algorithmic accountability and transparency"

36 Stankovich (2021), ITU: Emerging technology trends: Artificial intelligence and big data for development 4.0, https://www.itu.int/pub/D-TND-02

37 Haataja, M, van de Fliert, L and Rautio, P., (2020). Public AI Registers: Realising AI transparency and civic participation in government use of AI Saidot. Available at: https://ai.hel.fi/wp-content/uploads/White-Paper.pdf

38 Directive on Automated Decision-Making, https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

39 https://ainowinstitute.org/regulatingbiometrics.pdf

40 https://algorithmwatch.org/en/submission-to-un-report-on-right-to-privacy/

41 https://fra.europa.eu/en/publication/2022/bias-algorithm#publication-tab-1

42 https://www.whitehouse.gov/ostp/ai-bill-of-rights/

43 https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms

44 Ibid.

45 https://digital-strategy.ec.europa.eu/en/library/study-supporting-impact-assessment-ai-regulation

46 https://www.academia.edu/42244636/Algorithmic_transparency_and_the_right_to_explanation_Transparency_is_only_the_first_step

47 https://unicri.it/towards-responsible-artificial-intelligence-innovation

48 https://blog-idceurope.com/what-will-algorithmic-transparency-standards-mean-for-the-public-sector/

49 https://www.oecd-ilibrary.org/sites/96d49021-en/index.html?itemId=/content/component/96d49021-en

50 https://commonvoice.mozilla.org/en

51 https://www.cminds.co/eon-resilience-lab

52 https://www.weforum.org/reports/ai-procurement-in-a-box

53 https://aibusiness.com/verticals/brazil-pilots-ai-in-hospitals-metros-using-wef-best-practice-toolkit

54 https://publications.iadb.org/en/algorithmic-audit-decision-making-or-decision-support-systems; https://goblab.uai.cl/en/espanol-transparencia-algoritmica-en-el-sector-publico/

55 http://webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Screen_AW.pdf

56 https://karisma.org.co/

57 https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

58 City of Amsterdam Algorithm Register Beta. What is the Algorithm Register?, https://algoritmeregister.amsterdam.nl/en/ai-register/; City of Helsinki AI Register. What is AI Register? Available at: https://ai.hel.fi/en/ai-register/

59 http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf

SHAPING A MORE LIVABLE WORLD.

www.dai.com

@daiglobal