T20 Policy Brief

G20 BRASIL 2024

T20 Brasil 2024 Let's rethink the world

Task Force 05
INCLUSIVE DIGITAL TRANSFORMATION

Artificial Intelligence in Law Enforcement and Criminal Justice System: Transparency and Civil Participation as Core Issues in Promoting Human Rights

Cynthia Picolo, Executive Diretor, LAPIN (Brazil)

Kit Wash, Senior Staff Attorney, Electronic Frontier Foundation (United States)

Pablo Nunes, Coordinator , CESeC (Brazil)

Pedro Diogo Carvalho Monteiro, Surveillance Coordinator, LAPIN (Brazil)

Thalita Lima, Research Coordinator, CESeC (Brazil)

Veridiana Alimonti, Associate Director for Latin American Policy, Electronic Frontier Foundation (United States)

TF05

Abstract

In this policy brief, we analyze what to consider for mitigating risks of Artificial Intelligence (AI) use in law enforcement (LE), also assessing whether this technology is necessary or appropriate in each context. We focus on how improving transparency, public participation, and accountability are crucial to this task, especially involving the communities most affected by AI-based mass surveillance.

AI-based systems deployment in LE can affect fundamental rights like freedom of movement, speech, assembly, privacy, and data protection. A key concern is that AI in LE reproduces historical discrimination based, for instance, on race, gender, and class. Facial recognition technology (FRT) is an example of how systemic racism is ingrained in digital tools. Studies have shown that FRT has a high risk of error in non-white groups. It is also necessary to consider the impacts of AI-based decision-making technology on the presumption of innocence and due process when used in the criminal justice system.

Lack of transparency and civil participation surrounding government adoption of AI systems by LE aggravate these issues. Another concern is the limits imposed on public scrutiny and oversight over these tech initiatives, with obscurity becoming a feature of their implementation.

Policymaking on LE use of AI must ensure proper means of civil participation, especially by the groups mainly targeted by state mass surveillance. We relied on research evidence to elaborate on why governments must include these actors in policy decisions involving adopting AI for security purposes. We also discuss ways to enable civil oversight and its role in providing accountability for security agencies using AI. This subject is essential to the G20 agenda given the global nature of AI LE industry and its relevance for international cooperation.

Keywords: Artificial Intelligence, law enforcement, transparency, civil participation.

Diagnosis of the issue

The use of **Artificial Intelligence** systems in law enforcement has been a global concern in the last decade, including FRT in CCTV systems; predictive policing based on geographical criminal data; and recidivism risk evaluation in adjudicative decisions. The increasing adoption of these systems has been accompanied by mounting evidence of their problems and critical evaluations of their impacts.[1]

One crucial aspect cutting across these different implementations is their potential to perpetuate and exacerbate[2] discrimination against historically marginalized groups. This is because the development and deployment of these systems generally use databases biased by discriminatory patterns – in other words: garbage in, garbage out. As such, the patterns the AI models recognize and reproduce reflect racism and stigmatization based

[1] Hannah-Moffat, Kelly. "Actuarial Sentencing: An 'Unsettled' Proposition." Justice Quarterly 30, no. 2 (2013): 270-296; Hannah-Moffat, Kelly. "Algorithmic Risk Governance: Big Data Analytics, Race and Information Activism in Criminal Justice Debates." Theoretical Criminology 23, no. 4 (2019): 453-470; Brayne, Sarah. Predict and Surveil: Data, Discretion, and the Future of Policing. Oxford University Press, USA, 2020.

[2] Eubanks, Virginia. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press, 2018; Browne, Simone. Dark Matters: On the Surveillance of Blackness. Duke University Press, 2015; Benjamin, Ruha. "Race After Technology." In Social Theory Re-Wired, 405-415. Routledge, 2023; Crawford, Kate. The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, 2021.

on gender and economic disparities. The way LE agents and courts interact with these systems' outputs adds another layer to a pernicious dynamic establishing a discrimination feedback loop. This generates what scholars have called ‘algorithmic injustice’.

FRT performs particularly poorly in recognizing black people and ethnic minorities, as well as women, transgender, and nonbinary individuals. These groups face heightened risks of being misidentified and are disparately impacted by its use. Facial recognition can falsely accuse innocent people and make people into targets of unwarranted or dangerous retaliation.[3] There is a growing number of cases of wrongful arrests of innocent people.[4] Outdated databases of arrest warrants and wanted individuals worsen the problem.[5] The dissemination of FRT also raises serious concerns about enabling mass

[3] See an overview on how facial recognition technologies work and the problems its use entails at: https://sls.eff.org/technologies/face-recognition.

[4] Some of the occurrences of FRT errors, with wrongly arrests being done: Olhar Digital. Mulher é detida no Rio por erro em câmera de reconhecimento facial. July 7, 2019. See at:https://olhardigital.com.br/2019/07/10/seguranca/mulher-e-detida-no-rio-por-erro-em-camera-de-reconhecimento-facial/ CNN Business. A false facial recognition match sent this innocent Black man to jail. April 29, 2021. Johana Bhuiyan. First man wrongfully arrested because of facial recognition testifies as California weighs new bills. The Guardian, April 27, 2023. Kashmir Hill. Eight Months Pregnant and Arrested After False Facial Recognition Match. The New York Times. August 6, 2023.

[5] See more at https://www.law.georgetown.edu/privacy-technology-center/publications/a-forensic-without-the-science-face-recognition-in-u-s-criminal-investigations/. Accessed March 26, 2023.

surveillance, a disproportionate power that often fuels abusive practices in violation of fundamental rights.[6]

As FRT, predictive policing technology also contributes to arbitrary surveillance of people that unfair policing practices have historically targeted. Predictive policing systems use historical and sometimes real-time data to predict *when* and *where* a crime is most likely to occur or *wh*o is most likely to commit or be a victim of a crime.[7] The latter seeks to identify people; the former flags territories and communities. This technology is often a self-fulfilling prophecy since it tends to learn more about historical policing biases than about the actual patterns of crime incidence.[8]

The use of AI risk-assessment tools in the criminal justice system to assist highly consequential decisions like bail, sentencing, and paroles raises similar concerns. Such determinations, traditionally involving human and legal reasoning, are now increasingly

[6] See examples at James Vincent. NYPD used facial recognition software to track down Black Lives Matter activist. The Verge. August 18, 2020. Beatriz Busaniche. La Justicia define si vuelve el reconocimiento facial en las calles de Buenos Aires: ¿cómo auditar al vigilante artificial? elDiarioAR. February 19, 2024.

[7] Amnesty International (UK Section). Trapped in the Matrix: Secrecy, Stigma, and Bias in the Met’s Gangs Database. S.l., maio de 2018.

[8] Lum, Kristian, e William Isaac. "To Predict and Serve?" Significance 13, no. 5 (7 de outubro de 2016).

Richardson, Rachida, Jason M. Schultz, e Kate Crawford. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." New York University Law Review 94: 192-233.

mediated by AI.[9] One well-known example of how the adoption of these systems may have deleterious impacts on human rights is the criminal justice algorithm COMPAS - Correctional Offender Management Profiling for Alternative Sanction, a risk-assessment system used in the United States aimed at identifying defendants' tendency for recidivism. A study has shown that COMPAS reproduced the selectivity of the penal system against black defendants. The analysis found that black defendants who had not reoffended in two years were almost twice as likely to be wrongly classified as high-risk than white defendants.[10]

The use of these systems often shifts the traditional burden of proof away from law enforcement and courts, forcing targeted people to prove their innocence in the face of highly complex technologies and their opaque deployment by state agencies.

Our understanding is that these complications connect with the lack of civic participation and meaningful transparency on how government institutions develop, acquire, deploy, oversee, and evaluate these technologies. The adoption of technologies without participation and transparency threatens human rights, bolsters race-based discrimination, and undermines critical democratic foundations. As an example, the absence of specific data protection legislation in the context of LE and criminal prosecution in Brazil, as displayed in other Latin American countries, has favored a non-stop expansion in facial recognition usage in public security. The research collective *O*

[9] McKay, Carolyn. "Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making." *Current Issues in Criminal Justice* 32, no 1 (2020): 22-39

[10] Propublica. "How We Analyzed the Compas Recidivism Algorithm." Accessed March 25, 2024. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Panóptico pointed out the widespread use of FRT in the country, with 195 ongoing initiatives and 67 million Brazilians potentially surveilled by this technology.

G20 countries should carefully consider these trends and threats. Given the transnational nature of AI use in LE, with both companies' cross-border operations and security agencies' international cooperation, the G20 agenda must advance in a human rights-based approach to this issue. If digitalization is a central subject of the global economy, it's essential to discuss how it has affected law enforcement activities and its impacts on society. We believe that our recommendations can provide critical guidelines to address current concerns.

Recommendations

Faced with the present challenges, we group our recommendations into three central topics: (1) proper legal and institutional safeguards; (2) due diligence; and (3) meaningful transparency and participation. We detail each of them below.

(1) State use of AI systems in LE and criminal justice must be backed by proper legal and institutional safeguards	a) States must take seriously the need to adopt and effectively enforce, through independent, impartial, and well-resourced authorities, data protection legislation for the public and private sectors that complies with international human rights law, including safeguards, oversight, and remedies to effectively protect the rights to privacy and data protection, especially in LE; b) Deployment of AI systems for decision-making affecting rights in the LE context should be prohibited unless efficiency is proven and there is a legal basis for use through regulation, which must be democratically approved and informed by a previous and detailed impact assessment and public consultations.
(2) Conduct due diligence/ impact assessments before implementation and throughout AI systems' life cycle to decide	a) State adoption of AI-based technologies in LE context demands the previous conduction of thorough and multidisciplinary human rights impact assessment with inputs from experts and affected communities.

whether and how to adopt these systems	b) It also demands human oversight, and that States have in place the proper apparatus to prevent, address, and redress violations. This involves implementing solid monitoring and evaluation processes throughout the AI life cycle, including independent audits. c) On AI systems currently in use, promote joint efforts to suspend government use of remote biometric recognition technologies in public spaces, at least until authorities responsible can demonstrate compliance with privacy and data protection standards and the absence of significant accuracy issues and discriminatory impacts, and until all the recommendations set out in UNGA resolution A/HRC/44/24, paragraph 53 (j) (i–v) are implemented.
(3) Meaningful transparency and civic participation	a) It is crucial to effectively increase the transparency around government adoption of AI, including by appropriately informing the public and affected individuals and communities, as well as regularly providing data relevant for the public to assess systems' efficacy and impact on human rights; b) State actors should actively disclose meaningful information about their use of AI-based systems in law enforcement and criminal justice,

[redacted]	including details on developers and/or vendors, public budget, periodic auditing and impact assessment reports, performance metrics, technical documentation, and the decision-making flow; c) To ensure due process, AI systems employed in the LE and criminal justice impacting rights must meet interpretability and explainability standards. Defense experts must also have the ability to accede to detailed information needed to challenge the system.[11] d) State decision-making regarding the development, purchase, and implementation of AI technologies for security purposes should rely on effective participatory processes that include historically marginalized groups and affected communities. This involves conducting public hearings and meaningful consultation before and during their implementation. State actors, civil society, and academia must also build on shared knowledge

[11] See Karen Gullo. *Victory. New Jersey Court Rules Police Must Give Defendant the Facial Recognition Algorithms Used to Identify Him*. EFF. June 2, 2023.The court ruled that state prosecutors must turn over the defendant detailed information about the software, including how it works, source code, and its error rate.

[redacted]	about participation models to devise and establish more robust civic engagement and oversight.[12] e) States should commit to creating independent oversight bodies to review AI systems used for law enforcement and criminal justice, to assure algorithmic fairness and non-discrimination.

[12] Arnie Hintz et al. *Civic Participation in the Datafied Society: Towards Democratic Auditing?* Data Justice Lab. July 2022.

Scenario of Outcomes

The immediate benefit of approving and acting towards the implementation of these recommendations is reducing potential violation of human rights and risks of discrimination against targeted communities based on race, gender, ethnicity, and other difference markers.

Another important advantage of building on these recommendations to set a shared baseline approach among States would be the opportunity to develop international standards for government use of AI in law enforcement and criminal justice.

Given the international nature of the public security and surveillance market, the more States have structure and resources to adopt AI-based technologies following processes and priorities grounded on the protection of human rights, which are both transnationally shared, the more a human rights-based governance can become the prevailing model in such market. This would also contribute to avoiding deep global discrepancies in how States combat crime within and across borders. Global shared standards on which AI deployments are adequate within LE can mitigate situations where an implementation considered inherently harmful to human rights in one country is allowed in another despite raising the same risks and impacts. By agreeing to ban or suspend government deployment of harmful AI systems, State actors would improve the safety of their population and strengthen the common bases for their international cooperation in criminal matters.

The implementation of these recommendations also represents an economic gain for States. Governments would not only save the money invested in ineffective and controversial technology but also prevent financial losses resulting from their obligation to remedy violations caused by their use. The same logic applies to the deployment of AI-based technologies that, although not inherently problematic, are implemented without due consideration to aspects crucial for an effective and rights-respecting operation.

Having proper legal and institutional safeguards, upon which States can conduct solid impact assessments, that are informed by and integrate broader and meaningful transparency and participation mechanisms, allows a virtuous cycle to flourish. A virtuous cycle through which civil participation and oversight can provide critical elements for governments to evaluate and monitor the performance of AI-based policies and implementations in improving security while also identifying potential risks and violations of human rights. This outcome is deeply connected with increasing public spending efficiency, both by directing investments in technology that can truly advance public security and by preserving funds for other vital areas like education, healthcare, and culture.

Through our set of recommendations, State actors will be better positioned to foster a greater global and societal understanding around a lawful, accountable, and efficient deployment of AI. Doing so has also the potential to forward global shared values over topics like digital rights, data protection, and information security. A society that is more conscious about why and how State institutions adopt AI systems leads to an improvement in the relationship between those bodies and the community. This allows for more consistent feedback, monitoring, and evaluation of these policies and the role that AI plays in their success or failure.

References

Amnesty International (UK Section). *Trapped in the Matrix: Secrecy, Stigma, and Bias in the Met's Gangs Database*, 2018.

Benjamin, Ruha. "Race After Technology." In *Social Theory Re-Wired*, 405-415. Routledge, 2023.

Berk, Richard A. "Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement." *Annual Review of Criminology* 4 (2021): 209-237.

Bhuiyan, Johana. "First man wrongfully arrested because of facial recognition testifies as California weighs new bills." *The Guardian*, April 27, 2023.

Birhane, Abeba. "Algorithmic Injustice: A Relational Ethics Approach." *Patterns 2*, no. 2 (2021).

Brakel, Rosamunde. "Pre-emptive Big Data Surveillance and Its (Dis)empowering Consequences: The Case of Predictive Policing." In Exploring the Boundaries of Big Data, editado por B. V. D. Sloot, D. Broeders, e E. Schrijvers, Amsterdam: Amsterdam University Press, 2016.

Browne, Simone. *Dark Matters: On the Surveillance of Blackness*. Duke University Press, 2015.

Busanice, Beatriz. "La Justicia define si vuelve el reconocimiento facial en las calles de Buenos Aires: ¿cómo auditar al vigilante artificial?" *elDiarioAR*, February 19, 2024.

CNN Business. "A false facial recognition match sent this innocent Black man to jail." *CNN,* April 29, 2021.

Crawford, Kate. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.

Grinberg, Felipe et al. "Prisões por reconhecimento facial avançam pelo país, mas erros em série desafiam tecnologia de combate ao crime." *O Globo*, January 5, 2024.

Hannah-Moffat, Kelly. "Actuarial Sentencing: An 'Unsettled' Proposition." *Justice Quarterly 30*, no. 2 (2013): 270-296.

Hannah-Moffat, Kelly. "Algorithmic Risk Governance: Big Data Analytics, Race and Information Activism in Criminal Justice Debates." *Theoretical Criminology* 23, no. 4 (2019): 453-470.

Hintz, Arnie et al. *Civic Participation in the Datafied Society: Towards Democratic Auditing?* Data Justice Lab, 2022.

Jansen, Fieke. "Data Driven Policing in the Context of Europe." *Data Justice Lab*, ERC, Cardiff University, May 7, 2018.

Kashmir Hill. “Eight Months Pregnant and Arrested After False Facial Recognition Match.” *The New York Times*. August 6, 2023.

Lum, Kristian, e William Isaac. "To Predict and Serve?" *Significance* 13, no. 5, 2016.

McKay, Carolyn. "Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making." *Current Issues in Criminal Justice* 32, no. 1 (2020): 22-39.

Olhar Digital. “Mulher é detida no Rio por erro em câmera de reconhecimento facial.” July 7, 2019.

Propublica. "How We Analyzed the Compas Recidivism Algorithm." Accessed March 25, 2024. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Richardson, Rachida, Jason M. Schultz, e Kate Crawford. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *New York University Law Review* 94: 192-233.

Vincent, James. “NYPD used facial recognition software to track down Black Lives Matter activist.” *The Verge*, August 18, 2020.

G20 BRASIL 2024
BUILDING A JUST WORLD AND A SUSTAINABLE PLANET

T20 Brasil 2024
Let's rethink the world

Let's rethink the world

CEBRI
Rethink Tank

FUNDAÇÃO ALEXANDRE DE GUSMÃO

ipea
Instituto de Pesquisa Econômica Aplicada