T20 Policy Brief

G20 BRASIL 2024

T20 Brasil 2024 Let's rethink the world

Task Force 05

INCLUSIVE DIGITAL TRANSFORMATION

Dataset Discrimination in Government Surveillance: A Threat to Equality and Justice

Roberto Caparroz, Professor, FGV Direito SP (Brazil)

Chiara Battaglia Tonin, Student, Mackenzie University (Brazil)

Barbara Silverio Ferreira, Student, Glasgow (United Kingdom)

Aline Cruvinel, Student, Mackenzie University (Brazil)

Daniel Matos, Student, FGV EESP (Brazil)

Matheus Pereira, Student, Mackenzie University (Brazil)

Vitória Batista, Student, Mackenzie University (Brazil)

Luiza Balby, Student, FGV Direito SP (Brazil)

TF05

Abstract

The development of complex algorithms enabled the creation of artificial intelligence (AI) models that can act rationally and encode thousands of variables across millions of data points. These models have proven useful, including for the public sector. For example, machine learning has been used to predict European Court of Human Rights decisions; in Brazil, the Victor project is an AI system designed to identify general repercussions of pending cases in the Brazilian Federal Supreme Court to enhance analysis efficiency.

These tools have also been applied in government surveillance to identify individuals flagged as suspicious or dangerous, monitor crowds for potential threats, and help security staff locate lost children. The dataset discrimination debate is aligned with G20 priorities, as shown by the Sherpa Track's Digital Economy Working Group agenda. Also, some of its members can contribute with their own experience, as the European Union’s path about the AI Act.

The potential of AI-powered facial recognition in public spaces to enhance security and law enforcement is undeniable. Still, there are critical concerns about AI-based facial recognition in public spaces: discrimination and bias, lack of transparency and accountability, privacy violations, data protection, and cybersecurity. The G20 should establish a commission to investigate the ethical implications of AI and its associated machine learning and deep learning technologies, develop a model framework for using facial recognition in public spaces, outline common principles and minimum standards to guide national legislation, and launch a data governance initiative to promote harmonizing data protection standards.

Diagnosis of the issue

AI application in government surveillance is one of the various possible technology uses, but requests special attention. The massive use of facial recognition and machine learning in public spaces can be considered a tool to enhance security and law enforcement; although AI may undoubtedly contribute to public security under the above mentioned purposes and other justifications, there is also a risk of error, misuse, abuse, and biases, which must be considered.

The above mentioned risks are not restricted to academic papers and theoretical discussions; they have already materialized. The OECD AI Incidents Monitor, an initiative developed by the OECD.AI expert group on AI incidents with the support of the Patrick J. McGovern Foundation, has already mapped 9370[1] AI incidents[2]; the AI Incident

[1] The OECD AI Incidents Monitor may be consulted at https://oecd.ai/en/incidents?search_terms=%5B%5D&and_condition=false&from_date=2014-01-01&to_date=2024-03-29&properties_config=%7B%22principles%22:%5B%5D,%22industries%22:%5B%5D,%22harm_types%22:%5B%5D,%22harm_levels%22:%5B%5D,%22harmed_entities%22:%5B%5D%7D&only_threats=false&order_by=date&num_results=20. The number of AI incidents mapped was updated on March 29, 2024, and it may not be exact since an AI incident can be reported by one or more news articles covering the same event, according to OCDE.

[2] It is important to analyze the definition of AI Incident adopted by the initiative, which considers "an AI incident is an ev ent, circumstance or series of events where the development, use or malfunction of one or more AI systems directly or indirectly leads

Database, a broad catalog of harms reported by users worldwide, has already mapped 659[3] AI incidents[4] related to various entities from the private and public sectors.

Gender bias in AI is an example of a potential AI incident, and there are several ongoing investigations into its effects on different sectors and AI applications. Financial institutions are increasingly using AI for loan approvals, and these algorithms might perpetuate historical biases, where women are denied loans or offered worse rates compared to men with similar qualifications[5]. Also, facial recognition software has shown to be less accurate for certain skin classes (Joy; Gebru, 2018) - Joy and Gebru found out that "all classifiers tested performed best for lighter individuals and males overall. The classifiers performed worst for darker females".

---

to any of the following harms: (a) injury or damage to the health of a person or groups of people; (b) disruption of the management and operation of critical infrastructure; (c) violations of human rights or a breach of obligations under the applicable law intended to protect fundamental, labour and intellectual property rights; (d) damage to property, communities or the environment".

[3] The AI Incident database may be consulted at https://incidentdatabase.ai/. This number of AI incidents mapped was updated on March 29, 2024.

[4] The definition of AI incident adopted by AI Incident Database is "an alleged harm or near harm event to people, property, or the environment where an AI system is implicated".

[5] As an example, the US financial regulator has opened an investigation into claims Apple's credit card offered different credit limits for men and women ("Apple's 'sexist' Credit Card Investigated by US Regulator". November 10, 2019. https://www.bbc.com/news/business-50365609).

It is also important to consider that these biases originated in the data provided for AI training; the datasets reproduce the biases identified by contextual assumptions related to the data used for AI training. For example, to avoid the gender bias identified in facial recognition software, one of the proper measures is increasing phenotypic and demographic representation in face datasets and algorithmic evaluation (Joy; Gebru, 2018).

In this sense, AI and facial recognition in public spaces present opportunities and profound challenges. Responsible development and deployment demand a holistic approach prioritizing human rights, ethical considerations, and democratic control. By implementing the recommended policy frameworks and practical strategies, governments can navigate the labyrinth of AI and ensure its responsible use for a safer and more secure future without compromising individual liberties or fundamental rights. Only through thoughtful and proactive measures can we unlock the potential of this technology while safeguarding the values that underpin a just and equitable society.

Recommendations

The potential of AI-powered facial recognition in public spaces to enhance security and law enforcement is undeniable, but its use also poses substantial risks to fundamental rights. Robust policy frameworks and practical implementation strategies are imperative to enable its responsible use.

These measures are even more important in situations of conflict or radicalism, as in the recent case of the facial recognition system used by Israel, which catalogs the faces of unaware Palestinians without their consent and has resulted in the misidentification, abduction, and beatings of innocent Palestinians, including the poet Mosab Abu Toha[6].

In this sense, this document draws inspiration to chart a responsible course for AI and facial recognition in public spaces.

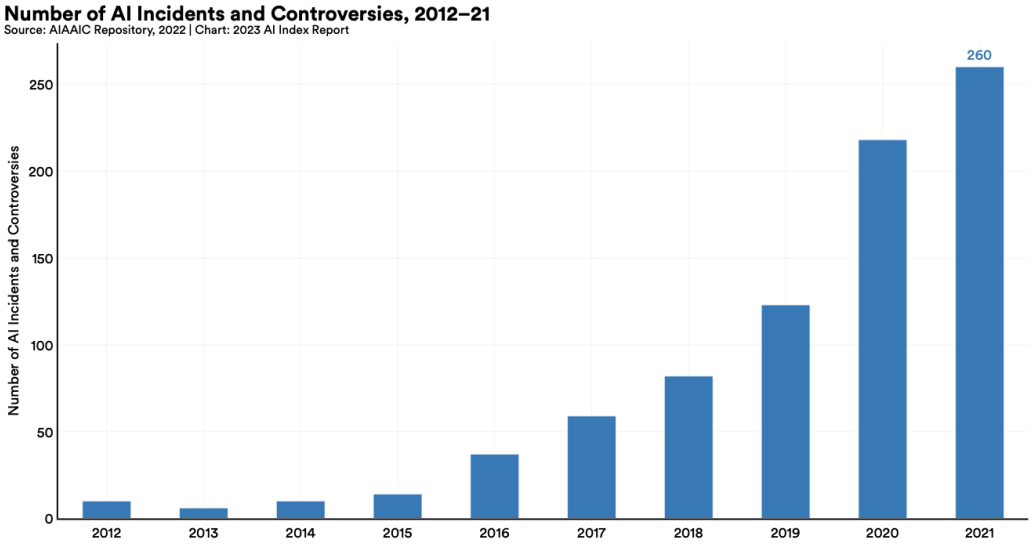
Key Concerns and Challenges:

- Discrimination and bias: Datasets biases embedded in facial recognition systems can disproportionately impact marginalized groups, leading to wrongful identification, profiling, and discriminatory practices.
- Lack of transparency and accountability: Opaque datasets, algorithms, and decision-making processes limit opportunities for correction in case of errors and facilitate misuse.

[6] https://www.nytimes.com/2024/03/27/technology/israel-facial-recognition-gaza.html

- Privacy violations: Massive data collection and retention are needed for facial recognition but raise concerns regarding constant surveillance and decreased individual privacy.
- Data protection and cybersecurity: Cyberattacks and unauthorized access to facial recognition databases pose significant risks, potentially compromising national security and individual safety.

The use of facial recognition systems on a large scale has the potential to increase the number of incidents involving artificial intelligence. The following graph shows that the number of AI incidents and controversies grew 26 times between 2012 and 2021, a 10-year interval.



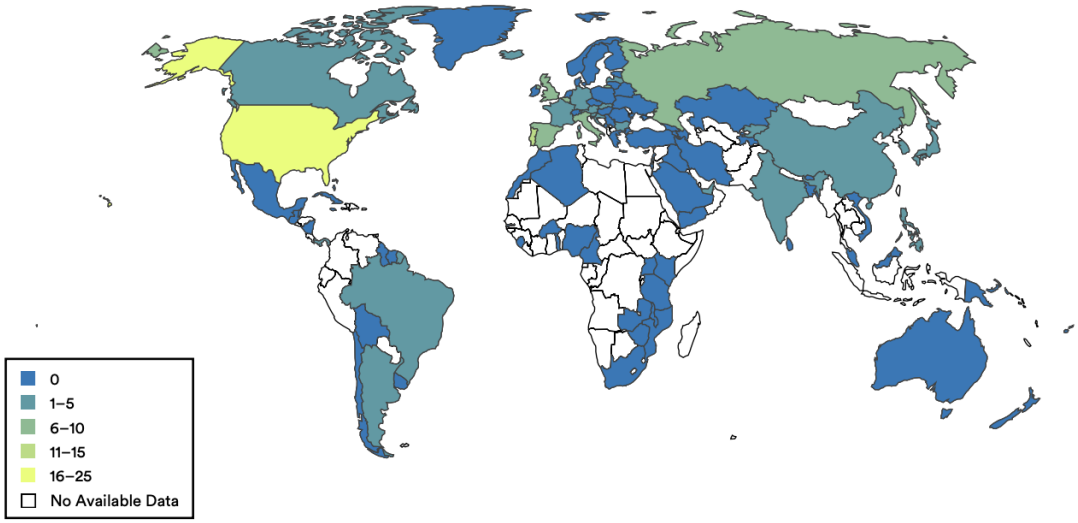
The problem is that AI systems have become ubiquitous, so much so that the number of incidents reported by the AIAAIC database reached 1409 at the beginning of 2024. This means an increase of more than five times in just two years. Therefore, the G20 countries must reflect on the regulation of AI models.

Policy Recommendations:

- Human oversight and control: High-stakes decisions like arrests or detentions must be previously assessed. Human oversight and judicial review are crucial to ensure accountability and prevent automated biases.
- Robust legal frameworks: Comprehensive legislation governing the use of facial recognition in public spaces should clearly define permissible uses, limitations, data protection requirements, and independent oversight mechanisms.
- Impact assessments and transparency: Entities must conduct ethical by-design processes, and impact assessments to evaluate potential discrimination and privacy risks.
- Data minimization and information security controls: Data collection and retention should be strictly limited to specific, legitimate purposes, adhering to the principles of necessity, proportionality, and data minimization.
- Public engagement and education: Broad public consultations and awareness campaigns are necessary to facilitate informed public debate and build trust in responsible AI governance. Educational initiatives should teach citizens about their rights and how to protect their privacy in the digital age.
- International cooperation and standardization: As facial recognition technology transcends national borders, international cooperation and harmonization of legal frameworks are crucial to prevent regulatory arbitrage and ensure global safeguards against misuse.

According to Maslej et al., policymakers' interest in AI is rising. The following graph shows that the number of bills related to "artificial intelligence" has grown more than sixfold in recent years.

Number of AI-Related Bills Passed Into Law by Country, 2016–22
Source: AI Index, 2022 | Chart: 2023 AI Index Report



Sparse rules show that countries are interested in regulating artificial intelligence models, but they may not be enough to guarantee fair and equitable treatment for all individuals. To this end, we believe it is essential to create a diverse commission, made up of various segments of society, with the aim of establishing frameworks and curation models for the large-scale development and use of datasets.

Scenario of outcomes

The G20 should establish a commission to investigate the ethical implications of AI and its associated machine learning and deep learning technologies. The commission and its work should complement the other G20 work on the broader international governance issues associated with AI. The commission should bring together academics specializing in AI, environment, and energy and industry and government representatives. A multidisciplinary approach will be essential for an effective response to what has so far been a relatively neglected aspect of technology in general and AI in particular.

The commission should be given the following tasks:

- Developing a model framework for using facial recognition in public spaces, outlining common principles and minimum standards to guide national legislation.
- Establish a panel of independent experts from various sectors to develop ethical guidelines for AI development and deployment, particularly in high-risk areas like facial recognition. Conduct independent assessments of national facial recognition programs, identify potential risks, and recommend best practices. Foster global dialogue and knowledge sharing on ethical AI governance.
- Launch a data governance initiative to promote harmonization of data protection standards and collaboration on developing secure and interoperable data infrastructures for international cooperation on facial recognition in specific circumstances, such as combating cross-border crime.
- Initiating a global public awareness campaign on AI and facial recognition, aiming to educate citizens about their rights and privacy implications of facial recognition technology, encourage public discourse and debate on the responsible use of AI in public

spaces, and promote transparency and accountability from governments and technology companies.

These actionable recommendations call for the G20 to lead by example, fostering international cooperation and promoting a responsible future for AI and facial recognition in public spaces. By leveraging their collective economic and political influence, the G20 can shape global norms and best practices, ensuring this powerful technology serves the public good while safeguarding individual rights and liberties.

Upon receipt of the report, the G20 should develop and publish an action plan to ensure the ethical and responsible use of AI-based facial recognition models in public spaces. This should include privacy and cybersecurity impacts and other AI-related global governance issues.

References

AIAAIC Repository. s. d. https://docs.google.com/spreadsheets/d/1Bn55B4xz21-_Rgdr8BBb2lt0n_4rzLGxFADMlVW0PYI/edit#gid=1051812323

Ammanath, Beena. *Trustworthy AI: a business guide for navigating trust and ethics in AI*. Hoboken, NJ: Wiley, 2022.

Angwin, Julia et al. "Machine Bias", *ProPublica*, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

BBC. "Apple's 'sexist' Credit Card Investigated by US Regulator". 2019. https://www.bbc.com/news/business-50365609.

Buolamwini, Joy and Gebru, Timnit. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. PMLR, 2018. https://proceedings.mlr.press/v81/buolamwini18a.html.

Christian, Brian. *The alignment problem: machine learning and human values*. New York, NY: W.W. Norton & Company, 2021.

Floridi, Luciano. *The ethics of artificial intelligence: principles, challenges, and opportunities*. New York: Oxford University Press, 2023.

Frenkel, Sheera. “Israel Deploys Expansive Facial Recognition Program in Gaza”. *The New York Times,* 2024. https://www.nytimes.com/2024/03/27/technology/israel-facial-recognition-gaza.html

Gary, Marcus and Davis, Ernest. *Rebooting AI: building Artificial Intelligence we can trust.* First Vintage Book edition ed. New York: Vintage Books, a division of Penguin Random House LLC, 2020.

Heaven, Will Douglas. “Predictive policing algorithms are racist. They need to be dismantled”. *MIT Technology Review*, 2020. https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

Liao, S. Matthew (Ed.). *Ethics of artificial intelligence*. New York, NY, United States of America: Oxford University Publication, 2020.

Maslej, Nestor et al. “The AI Index 2023 Annual Report,” *AI Index Steering Committee*, Institute for Human-Centered AI, Stanford University, Stanford, CA, 2023.

McGregor, Sean. “Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database”. In *Proceedings of the Thirty-Third Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-21)*. Virtual Conference, 2021.

OECD. “AIM: The OECD AI Incidents Monitor, an Evidence Base for Trustworthy AI”. s. d. Access in March 29, 2024. https://oecd.ai/en/incidents.

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