T20 Policy Brief

G20 BRASIL 2024

T20 Brasil 2024 Let's rethink the world

Task Force 05

INCLUSIVE DIGITAL TRANSFORMATION

Auditing Algorithms as Way to Prevent Algorithmic Bias

Cecilia Danesi, Researcher, Institute for European Studies and Human Rights at the Pontifical University of Salamanca (Spain)

Mario Torres Jarrín, Director, Institute of European Studies and Human Rights at the Pontifical University of Salamanca (Spain)

TF05

Abstract

We have no doubt that artificial intelligence became an essential part of the development of modern societies. There are many areas, crucial for everyday life, which could no longer function without algorithms. However, artificial intelligence has a main danger which could be sum up in one term: "algorithmic bias". This means that our prejudices are transferred to AI systems and the result is the amplification and consolidation of social gaps. That is the direct violation of Human Rights and the constant distance from the SDGs.

For that reason, G20 Governments must construct a coercive ethical framework for AI. This Ethics governance of AI (EGAI) should include different main backbones, such as 1) create an interdisciplinary expert committee, 2) design an AI strategic plan, 3) establish the mandatory supervision of high-risk algorithms, 4) elaborate studies that identify disruptive technologies and indicates their impact on our societies, 5) establish permanent dialogues G20-Big Tech Companies, 6) encourage to T20 System to carry out joint research on the impact of disruptive technologies and 7) promote reform in education system at all levels (primary, secondary, higher education and vocational training) that include learning about disruptive technologies.

Keywords: auditing algorithms, algorithmic bias, artificial intelligence, human rights, ethics governance of artificial intelligence.

Diagnosis of the issue

Artificial intelligence became an essential part of the development of modern societies. There are many areas, crucial for everyday life, which could no longer function without algorithms. One of the most outstanding cases are digital services that, considering the definition provided by a regional integration organization as the European Union (EU), we refer to the Digital Services Act which includes online intermediaries and platforms such as online marketplaces, social networks, content-sharing platforms, app stores, and online travel and accommodation platforms. Moreover, GenAI arrived not only to popularize the use of AI, but also to bring new risks, e.g. the access to the creation of false content, contributing to the amplification of hate speech and disinformation with a huge impact into democracy and Human Rights. It is undoubtedly, an important initiative in terms of regulation on digital affairs but being regional in scope it is not enough since regulation at a global level is required.

On the one hand, if we consider that the main Big Tech Companies are from the United States and China, besides these countries together with Canada, Brazil, India, United Kingdom, and EU looking for regulate emerging technologies, mainly, the Artificial Intelligence, then the G20 is a more suitable forum to promote initiatives that seeks to create international norms and development of global data governance standards. On the other hand, the "Big Tech Companies" are that control and management the cyberspace, these companies lead the "Industry 4.0" sector and accumulate market capitalization exceeds to the GDP of the most bigger economies around the world, what make in the new geopolitical actors in the international arena, its products (goods and services) and investments decisions have implications ion the international trade, global governance, definitely in the world order system. In this sense, the debate on auditing algorithms as

way to prevent algorithmic bias should be led by the G20. It is necessary to implement a G20 Tech Diplomacy (Riordan & Torres Jarrín 2021, Torres Jarrín & Riordan 2023).

The main danger of the use of AI could be sum up in one term: "algorithmic bias". This means that our prejudices are transferred to AI and the result is the amplification and consolidation of social gaps. That is the direct violation of Human Rights and the constant distance from the SDGs. Unfortunately, we can find several examples of algorithmic bias: Amazon system which discriminated female candidates and chose men instead, Google photos algorithm that classified African American persons as gorillas, IDEMIA'S facial recognition algorithm biased against black women, or social media filters that reinforce stereotypes, just to name a few among millions. GenAI is not out of these problems: a study of more than 5.000 images created with Stable Diffusion found that it has gender racial biases to extremes (Bloomberg). It shows a worse scenario than in the physical world (Danesi 2022).

Addressing "algorithmic biases" faces with several and huge issues. On the one hand, the ubiquity of artificial intelligence and, in turn, its imperceptibility. Nowadays, we have no doubts that AI is everywhere (specially with the launched of ChatGPT), but we cannot detect its presence and even less identify if it is biased. This happens mainly because there is not any obligation to inform that people is interacting or influencing by AI and neither that companies must transparent that how its algorithms work. Knowing how algorithms are used and specially their objectives and variables will give us useful information to detect biases or inappropriate uses. On the other, once transparency and explainability is mandatory, we can evaluate algorithms to identifies biases, discrimination, unfavorable treatment of vulnerable/minority groups, non-representative or reliable datasets, among others.

Different countries and international organizations are working on building an Ethics Governance of AI. This consists of creating a huge umbrella of actions and initiatives to guarantee an ethical use of AI. Through a strategic plan of AI, Governments must stablish their priorities and "steps forwards" with an interdisciplinary (health, law, education, finance, etc.), multisectoral (academia, civil associations, public and private sector, etc.) and gender equality perspective (Mota Sanchez 2023).

Nevertheless, most of the countries limit their efforts in ethics principles, which without the force of law, have no sense and end up being a dead letter. We can find some exceptions. The European Union Parliament has approved the AI Act; the first law in the world that regulates artificial intelligence in a comprehensive and preventive way. The rule classifies AI systems by risk: unacceptable, high, limited, and minimal. It also stablishes norms for general AI models and creates the AI Office inside the European Commission. With respect to our point of analysis, the Act provides a list of requirements for high-risk systems (which includes the examination of possible biases that may affect the health and safety of people, negatively affect fundamental rights, or give rise to any type of discrimination prohibited by Union Law, art. 10) and the obligation to do a fundamental rights impact assessment.

Apart from EU regulation, we can find local laws that pursue the same objective. NYC's automated employment decision tool law stablishes that employers who use AI in hiring or for deciding promotions must tell candidates they are doing so and must do annual independent audits to detect and prevent algorithmic biases (Aránguiz Villagrán 2022). Moreover, some countries are creating agencies for the supervision of Artificial Intelligence, such as the case of Spain.

Finally, the idea of building an Ethics governance of AI (EGAI) which includes the algorithmic audit of high-risk AI systems (among other actions which will be mentioned

below), is also aligned with several of the issue notes of the G20. For example, one of the fundamental pilar of the EGAI is the respect for the environment. AI could be an ally for sustainability, helping to find “Innovative Perspectives on Fuels Sustainable” (see Brazil-Concept note) but, at the same time, a big polluter. Furthermore, artificial intelligence could contribute for inclusive sustainable development and inequalities reduction (see Digital economy issue note) and to close the gender gap and protect women’s rights in order to achieve SDG 5 (see Empowerment of Women note)

Recommendations

Considering that worrying scenario, we strongly believe that democratic societies that respect their rights must construct a coercive ethical framework for AI. The OECD Principles on Artificial Intelligence, Recommendation on the Ethics of Artificial Intelligence adopted by UNESCO, or the EU AI Act can be served as guidelines to promote an Ethics Governance of Artificial Intelligence (EGAI) at global level. This EGAI should include different main backbones:

1. Create an interdisciplinary expert committee as an advisory body to assist the public sector and issue best practice guides for the ethical development of AI.
2. Design an AI strategic plan to specify the priorities and main steps of each government.
3. Establish the mandatory supervision of high-risk algorithms. We are referring to AI systems that could cause a huge harm in people's lives, for example, discriminating or creating an unfavorable treatment.
4. Elaborate studies that identify disruptive technologies and indicates their impact on our societies. Based on this study, a multi-stakeholder international conference (government, private sector, academia, and civil society) is convened to establish an international treaty on the regulation of technologies disruptive.
5. Establish permanent dialogues G20-Big Tech Companies. The G20 should appoint a Tech Ambassador to develop relationships with Big Tech Companies and coordinate the organization of the international conference on disruptive technologies.
6. Encourage to T20 System to carry out joint research on the impact of disruptive technologies in all areas of societies as legal, economic, social, cultural and environmental.

7. Promote reform in education system at all levels (primary, secondary, higher education and vocational training) that include learning about disruptive technologies.

Scenario of Outcomes

G20 should recommend a comprehensive initiative of Ethics Governance of Artificial Intelligence (EGAI), which has to include a mandatory algorithmic audit for high-risk AI systems. This has to be part of a strategic and comprehensive AI Ethical Governance plan that not only seeks to prevent algorithmic biases through audits (and education on the subject) but also to encourage innovation and investment in the field from an ethical approach.

This set of measures will allow G20 members to obtain the following results:

- Predictability and legal certainty.
- Promotion of innovation and investment.
- Legal and geopolitical strategy harmonization among G20 members.
- Construction of an artificial intelligence ecosystem that respects the legal framework, especially Human Rights.
- Global positioning of G20 members with respect to Big Tech Companies.

In conclusion, the G20 leadership on Ethics Governance of Artificial Intelligence and establishing of a permanent dialogue between G20-Big Tech Companies through a G20 Tech Diplomacy can contribute to G20 countries to led the Global Tech Governance future.

References

Aránguiz Villagrán, M. Auditoría algorítmica para sistemas de toma o soporte de decisiones, BID, 2022.

Castellanos J. et al. Transparencia y explicabilidad de la inteligencia artificial, Tirant lo Blanch, Valencia, 2022.

Danesi, C. Struggling against algorithmic bias: civil liability and other legal remedies. Cacucci Editore, 2022.

Danesi, C. The empire of algorithms, Galerna, 2022.

Mota Sanchez, E. & Herrera Expósito, E. Algorithmic audit in artificial intelligence in the public sector, Revista Proyecciones, nº 17, 2023.

Riordan, Shaun & Torres Jarrín, Mario. "A G20 Tech Diplomacy", Policy Brief Task Force 8 Multilateralism and Global Governance, G20 Presidency-Italy September 2021.

Torres Jarrín, Mario and Riordan, Shaun. *Science Diplomacy, Cyberdiplomacy and Techplomacy in EU-LAC relations.* Cham: Springer Nature Switzerland AG, 2023.

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