



Task Force 05

INCLUSIVE DIGITAL TRANSFORMATION

AI Technologies: Algorithmic Monocultures, Arbitrariness, and Global Divides

Claudio Mayrink Verdun, Research Scholar, School of Engineering and Applied Science, Harvard University (USA)

Camila Akemi Tsuzuki, Operations Coordinator, Instituto Vero (Brazil)

Caio Vieira Machado, PhD Student, Centre for Socio-Legal Studies, University of Oxford, UK and Universidade de São Paulo Law School (Brazil)

Juan Felipe Gomez, PhD Student, Department of Physics, Harvard University (USA)

Lucas Monteiro Paes, PhD Student, School of Engineering and Applied Science, Harvard University (USA)

Flavio P. Calmon, Assistant Professor, School of Engineering and Applied Science, Harvard University (USA)



Abstract

Artificial intelligence (AI) technologies are becoming increasingly complex, powerful, and opaque. The multi-million dollar cost of developing AI models has concentrated state-of-the-art AI production within entities in the Global North. The majority of AI systems are designed by a limited set of actors, leading to a uniformity that fails to reflect the pluralistic nature of societies, potentially overlooking or misinterpreting the needs and values of Global South populations [10]. This centralization fosters an "algorithmic monoculture:" a common set of AI models is deployed across multiple decision-makers. This monoculture paves the way for the emergence of "*Algorithmic Leviathans*" [3,13] – where AI exercises disproportionate control over access to resources and the exercise of rights and can exacerbate inequalities and introduce new forms of discrimination. This policy brief recommends an ethical governance framework for mitigating two harms of Algorithmic Leviathans: arbitrariness and bias.

1) Arbitrariness manifests when seemingly innocuous decisions during development of AI models lead to unexpected and detrimental individual and collective outcomes. Arbitrariness needs to be addressed by specific mechanisms of transparency as well as standardized procedures for evaluating and reporting potential harms, even at the early stages of AI development. 2) Bias and discrimination occur when an AI system's performance varies across legally protected populations and groups. These discriminatory effects produce inequalities at scale in algorithmic Leviathans and monoculture contexts. Such effects are aggravated because the Global North does not consider the Global South's legal and demographic contexts. To remediate arbitrariness and bias, we recommend systematic processes for human appeal and review. These safeguards are essential for AI to foster social development, mitigate inequality, and advance social justice.

Keywords: Algorithmic Monocultures, Predictive Multiplicity, Fairness

Diagnosis of the Issue

This brief offers recommendations for the G20 economies to promote diversity in training data, mitigate algorithmic biases, and ensure that AI technologies serve the interests of all global communities equitably. In particular, we discuss AI monocultures, which upscales distortion, and algorithmic arbitrariness, which can lead to unpredictable individual predictions from AI algorithms. These issues, in turn, influence the decisions taken by those deploying algorithms, amplifying systemic injustices at local and global levels. Furthermore, the division between the Global North and the Global South intensifies these risks as there is a growing concern that the benefits and risks of AI deployment may not be equitably distributed worldwide.

The Rise of AI Monocultures

AI monocultures, characterized by the dominance of a singular AI system or technology across industries, raise concerns akin to those observed in agricultural monocultures where a single disease can devastate entire crops [1,2]. For example, powerful Large Language Models are produced by a handful of companies and cost billions of dollars to produce. This uniformity poses risks of systemic failures, where vulnerabilities or flaws in the dominant approach could lead to widespread disruptions across multiple applications. Moreover, it can narrow innovation by directing resources towards a single technology, discouraging exploration of alternative AI systems that may offer solutions to unaddressed problems. Additionally, AI monocultures exacerbate ethical concerns, as biases inherited from training data can be entrenched across various applications, undermining fairness and ethical decision-making. The concentration of market dominance within certain entities also raises issues of competition suppression

and control over AI's societal application, necessitating comprehensive consideration of regulatory and economic dimensions for a resilient AI ecosystem.

The global dynamics of AI raise concerns about monocultures, with tools from the Global North potentially dominating the Global South, amplifying biases, controlling resources, and failing to represent diversity. Governance frameworks may perpetuate inequalities influenced by the Global North. The G20 must address global AI challenges associated with deployment and regulation without accelerating inequality.

Monocultures and the Risks of "Algorithmic Leviathans"

The concept of Algorithmic Leviathans denotes the substantial shift towards the utilization of algorithms and automated decision-making systems across high-stakes domains like the judiciary, healthcare, finance, and national security, characterized by centralized control and minimal human oversight [3]. This reliance on extensive data collection to fuel precise decision-making raises concerns about opacity, accountability, and privacy, posing risks such as erosion of individual autonomy and perpetuation of biases. In various sectors, as those mentioned above, the deployment of AI algorithms introduces potential biases and discriminatory outcomes [4], disproportionately affecting marginalized groups and exacerbating systemic inequalities [12]. The complexity and opacity of algorithmic decision-making challenge accountability and ethical oversight, posing challenges in discerning the rationale behind specific decisions and ensuring fairness in outcomes. This underscores the urgent need for robust ethical frameworks and transparency measures to address the risks associated with Algorithmic Leviathans and mitigate their adverse impacts on society such as threats to freedom of expression [9].

Conceptualizing and identifying algorithmic arbitrariness

We call the G20 economies' attention to groundbreaking research in the AI community on algorithmic arbitrariness. The training processes used to create AI models are incredibly complex. Opaque choices made during model development can lead to competing models with similar average performance on a given task but conflicting individual predictions [8]. This phenomenon, known as predictive multiplicity [5], results in arbitrary outcomes for users when the choice among competing models lacks transparency and consistency.

Predictive multiplicity arises, in part, due to an over-emphasis on measuring model performance solely based on average accuracy while neglecting critical criteria such as accountability, fairness, and transparency. For instance, two AI models used in criminal justice systems can have comparable average performance, yet disagree on whether an individual defendant's likelihood of failing to appear in court [12]. Arbitrarily choosing between these models leads to an arbitrary outcome for that defendant. Selecting an AI model becomes challenging when alternative models generate more favorable predictions for some individuals, highlighting the need for robust regulations and clear guidelines to ensure fair and transparent decision-making processes. Moreover, biased or unrepresentative training data used to develop AI models can exacerbate multiplicity and systematically disadvantage marginalized population groups [5, 9]. Without transparency, complex AI models resemble slot machines, with outcomes and predictions for vulnerable users determined from random, opaque, and unjustified choices made during their development. Data quality and selection play a crucial role, where biased or unrepresentative training data can skew algorithmic decisions, systematically disadvantaging certain groups and increasing arbitrariness [6, 9]. All this together shows that selecting a predictive model can be challenging, especially when alternative models

yield more favorable predictions for certain individuals. Solving this multiplicity issue, which remains an open problem in AI research, will help promote accountability and transparency in AI.

Recommendations

1. Managing Arbitrariness in Algorithmic Systems

Establishing ethical guidelines for AI development and usage is vital for G20 economies to mitigate arbitrary decisions from AI algorithms. These guidelines ensure adherence to fairness, transparency, and accountability principles throughout the AI lifecycle, promoting responsible development. Implementing measures for bias detection and mitigation is essential, where diverse, representative datasets and fairness-aware learning techniques can reduce bias influence. Such policy interventions foster a more ethical AI ecosystem, enhancing public trust in AI technologies.

We urge G20 economies to prioritize discussions on arbitrariness in AI agendas. Addressing arbitrariness requires a comprehensive approach that tackles its multifaceted causes. First, it is important to recognize the dominance of AIs developed in the Global North and the effects that arbitrariness can carry in the North-to-South dynamics. Second, countries need to deploy dedicated mechanisms for transparency, accountability, and fairness. Enhancing data quality, algorithm transparency, and ethical design are crucial, requiring adapted regulatory frameworks for effective governance. These solutions must operate globally to ensure AI contributes positively to society while upholding fairness and accountability and should not be restricted to jurisdictions.

2. Technical Solutions

Innovative technical measures are pivotal in addressing risks linked to AI systems. A recent avenue of research aimed at mitigating multiplicity effects involves using selective ensembles to reduce inconsistencies across AI algorithms [7,11]. These techniques identify and flag data points susceptible to inconsistency, providing a method for meticulously addressing their classification. Employing an ensemble of models can significantly mitigate inconsistencies across AI algorithms, addressing concerns about arbitrariness. By combining multiple models, each trained on diverse datasets or employing different methodologies, an ensemble approach reduces the risk of individual biases or arbitrary decisions dominating the overall system. This helps balance out discrepancies and promotes more robust and reliable predictions. We mention below a few avenues for G20 economies to tackle these issues:

- One approach could involve the development of **standardized protocols and frameworks for data exchange and model interoperability, ensuring compatibility and consistency across borders**. G20 economies can incentivize international collaborations and partnerships through funding initiatives, research grants, and joint projects aimed at pooling resources and expertise for developing shared AI solutions. By leveraging their influence and resources, G20 countries can foster a conducive environment for collaborative AI innovation, driving progress toward more equitable and globally beneficial outcomes.

- G20 economies should prioritize the development of **audit trails, ethical guardrails, and interpretability to promote ethical governance of AI**. These solutions safeguard fundamental rights like privacy and equal treatment and should align with societal interests guided by the state. By enhancing security, ensuring ethical compliance, and preventing misuse, these technical advancements contribute to transparency and

accountability in AI decision-making processes. For instance, audit trails and ethical guardrails increase transparency and accountability, while privacy-enhancing technologies protect user data, fostering trust in AI applications. Improving interpretability ensures that AI decisions are understandable and trusted, while robust data management minimizes biases and inaccuracies, preventing the perpetuation of societal inequities.

3. Policy and Law

On the policy and legal front, establishing clear requirements and frameworks for the responsible development and deployment of AI is fundamental. Such solutions can take a number of forms:

- **Incentives for developers:** disclosure of potential risks and engagement in responsible AI use can cultivate an environment where ethical considerations are prioritized from the outset.
- **Mandatory risk assessments:** before AI systems are released, ensure that potential harms are identified and mitigated early on. The designation of cross-country chief AI officers responsible for the test and deployment of public models could accomplish this.
- **User manuals:** Requiring AI developers to clearly define if the intended uses of their systems fosters transparency and prevents the application of AI in inappropriate contexts.
- **Quantitative metrics:** Broadening the metrics for evaluating AI systems beyond technical accuracy to include fairness, arbitrariness, transparency, and societal impact.
- **Data Protection and Legal Accountability:** Strengthening data protection regulations to include provisions for AI risk assessments and requiring companies to

justify their model decisions and risk management practices from a legal perspective. This aims to ensure that AI systems are developed and deployed responsibly, with a clear understanding of their potential impacts on individuals and society.

- **Awareness:** Promoting education and awareness initiatives to increase understanding of AI technologies and their potential impacts can empower individuals and organizations to identify and address arbitrary decisions. This could involve public education campaigns, training programs for AI developers, and resources for policymakers to stay informed about AI-related issues.

The development of robust auditing frameworks facilitates continuous oversight and accountability, ensuring AI systems remain aligned with ethical guidelines and regulatory standards over time. Comprehensive regulation addressing every phase of an AI system's lifecycle ensures that ethical considerations are integrated from the initial design to deployment and operation, fostering a more responsible and beneficial application of AI technologies.

4. Addressing the issue of Arbitrariness

Along with the technical and policy solutions proposed above, this brief suggests several technically informed solutions to enhance AI systems' transparency, fairness, and accountability to counteract arbitrariness. These include:

- **Developing New Arbitrariness Metrics:** Beyond assessing impact, new metrics that can specifically identify and quantify arbitrariness and its effects are needed. This approach helps in understanding and mitigating arbitrary decisions within AI systems.
- **Cost-efficient Arbitrariness Mitigation:** Finding cost-effective methods to assess the arbitrariness of AI models is crucial. This enables more widespread and routine

evaluations of AI systems, ensuring they function as intended without imposing undue burdens.

- **Global Discrepancies:** It is crucial to consider the impact of decisions made by AI systems developed in the Global North when applied in the Global South. Arbitrariness has disparate effects on the population, and its effects may be exacerbated on marginalized populations [4, 9] and developing economies. Policies and regulations should be designed to accommodate these differences, ensuring AI applications are equitable and do not perpetuate existing biases or inequalities.

The Global Challenge of Fairness in AI: Scenarios of Outcomes

Aware of the advances in the research of AI fairness and arbitrariness, it is necessary that G20 countries foster a collaborative cross-country environment for combining AI models trained on diverse datasets from different nations. As effects have the potential to affect billions of people across various countries, as we have demonstrated with our research in content moderation [10], the research and accountability measures need to be as wide-reaching as the AI service itself so people can use these tools in their own jurisdictions. Establishing public policies that facilitate data sharing and model collaboration can enable the creation of more robust and globally inclusive AI systems.

Economic disparities are also pronounced, as the technologically advanced nations producing AI capture a significant share of the economic benefits. This dynamic positions countries in the Global South primarily as consumers, rather than producers, of AI technology, hindering their ability to leverage AI for local development and growth. The regulatory landscape presents additional hurdles for the Global South, where limited resources, expertise, and suitable legal frameworks make it challenging to govern foreign

AI technologies effectively. This vulnerability prevents countries from protecting their citizens adequately or holding foreign entities accountable.

Lastly, dependency on AI technologies from the Global North can foster technological dependence, limiting the Global South's capacity to develop independent technological solutions. This dependence jeopardizes technological sovereignty and the ability to customize AI applications to meet local needs and challenges. We point to a few important scenarios that G20 economies need to take into account.

First, promoting fairness in AI and reducing inequalities requires collaboration between countries that are at different stages of the AI industry and benefit differently from this sector.

The fairness issues arising from the production of AI in the Global North and impacting the Global South are multifaceted. One significant concern is bias and representation in AI systems, where datasets predominantly reflect the demographics and languages of the Global North. This skew results in AI applications, like facial recognition and language processing systems, performing inadequately for individuals from the Global South. Such biases not only lead to discriminatory outcomes but further marginalize disadvantaged groups, exacerbating existing disparities.

Data exploitation is a critical issue where personal data from the Global South is harvested and used by companies in the Global North without fair compensation or adherence to privacy and ethical standards. This "data colonialism" undermines community autonomy and contributes to a digital divide.

Algorithms lacking contextual understanding can disregard human-relevant factors, enhancing the perception of arbitrariness. Implementation choices can introduce subjectivity, and societal biases embedded in algorithms can lead to decisions that reflect existing prejudices, adding another layer of arbitrariness.

Although urgent and necessary, the aforementioned measures possess certain risks, such as (i.) human oversight limitations: while human oversight is essential, relying solely on human intervention may not be feasible or scalable, particularly where AI systems process vast amounts of data in real-time. (ii.) overregulation: excessive regulatory frameworks could stifle innovation and hinder the development and deployment of AI technologies. Striking a balance between regulation and innovation is crucial to avoid impeding technological progress and, finally, (iii.) compliance burden since companies operating in multiple countries may face compliance challenges as they navigate different legal requirements and ethical standards. Adhering to disparate regulations increases administrative burden and compliance costs, potentially hindering innovation and global collaboration.

Accommodating these national differences around AI might be challenging and costly at first, but it will usher profound economic and social transformations. This is an important opportunity and the right timing for G20 economies to collaborate and promote joint social and economic development.

References

- Jon Kleinberg and Manish Raghavan. "Algorithmic monoculture and social welfare." *Proceedings of the National Academy of Sciences* 118.22 (2021): e2018340118.
- Rishi Bommasani, Kathleen A. Creel, Ananya Kumar, Dan Jurafsky, and Percy S. Liang. "Picking on the same person: Does algorithmic monoculture lead to outcome homogenization?." *Advances in Neural Information Processing Systems* 35 (2022): 3663-3678.
- Pascal König."Dissecting the algorithmic leviathan: On the socio-political anatomy of algorithmic governance." *Philosophy & Technology* 33.3 (2020): 467-485.
- Julia Dressel, and Hany Farid. "The accuracy, fairness, and limits of predicting recidivism." *Science advances* 4.1 (2018): eaao5580.
- Charles Marx, Flavio Calmon, and Berk Ustun. "Predictive multiplicity in classification." *International Conference on Machine Learning*. PMLR (2020).
- Anna Meyer, Aws Albarghouthi, and Loris D'Antoni. "The dataset multiplicity problem: How unreliable data impacts predictions." *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. (2023).
- Emily Black, Klas Leino, and Matt Fredrikson. "Selective Ensembles for Consistent Predictions." *International Conference on Learning Representations*. (2021).
- Emily Black, Manish Raghavan, and Solon Barocas. "Model multiplicity: Opportunities, concerns, and solutions." *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. (2022).
- Juan Felipe Gomez, Caio Vieira Machado, Lucas Monteiro Paes, and Flavio P. Calmon. "Algorithmic Arbitrariness in Content Moderation." *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. (2024).

Marie-Therese Png. "At the tensions of south and north: Critical roles of global south stakeholders in AI governance." *In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1434-1445. (2022).

Carol Long, Hsiang Hsu, Wael Alghamdi, and Flavio Calmon. "Individual Arbitrariness and Group Fairness." *Advances in Neural Information Processing Systems* 36 (2024).

Emily Black, John Logan Koepke, Pauline Kim, Solon Barocas, and Mingwei Hsu. "Less Discriminatory Algorithms." *Available at SSRN* (2023).

Kathleen Creel and Deborah Hellman. "The algorithmic leviathan: Arbitrariness, fairness, and opportunity in algorithmic decision-making systems." *Canadian Journal of Philosophy* 52.1 (2022): 26-43.



Let's **rethink** the world

