



Call for Papers on Global AI Governance

Office of the UN Secretary General's Envoy on Technology

Agustina Del Campo

CELE

adelcampoiso@gmail.com

September 28, 2023

Introduction

The Center for Studies on Freedom of Expression and Access to Information (CELE) is a global research Center based in Argentina and affiliated with Universidad de Palermo. The Center works to promote and enhance the protection of freedom of expression through cutting-edge research capable of shaping and changing public debate on key policy issues, and capacity building.

Over the past few years there has been increasing concern over AI and its impacts on society. Efforts to address these challenges have mostly focused on the technology itself. We believe that the UN should shift the focus and reframe this conversation within a rights-based approach. This submission will first make a short summary of current approaches and then propose that the UN adopts a strategic human rights based approach to AI governance.

Current approaches to AI governance

There have been several initiatives around AI governance in the past few years. They differ in nature and scope, some being international recom-

mentations, others bills of law or enacted regulation. Key among them, there were the Global Partnership on AI (GPAI) (2018), the Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on AI (2018); the OECD AI Recommendations (2019); the UNESCO recommendations on the ethics of AI (2020); the EU AI Act of 2021; Canada's The Artificial Intelligence and Data Act (2022); The AI Bill of Rights (2022); The G7 Hiroshima AI process (2023); and China's Interim Administrative Measures for Generative Artificial Intelligence Services (2023). Most of them have one thing in common: they approach AI as a singular cohesive technology to be governed independently of other the issues it raises. David Kaye's report may be cited as an exception.

The Report of the Special Rapporteur in 2018 was among the first attempts to address the challenges and opportunities raised by AI for freedom of expression online. The report is specific to freedom of expression and even more particularly to the use of AI within social media and search engines. The GPAI is an initiative created by the G7 also in 2018, multistakeholder, which gathers States, civil society, industry and academics to join forces to address the challenges brought about by AI. Unlike the report, it is specific to AI as a technology rather than specific to the impacts of AI on a concrete human right. They do provide for working groups on responsible AI, data governance, future of work, and innovation and commercialization, which allows for more nuanced approaches to the technology itself. They [adhere to the OECD guidelines and recommendations](#) and the OECD hosts the secretariat of the GPAI.

The OECD Recommendations take a broad, industry-oriented approach to AI mandating transparency and explainability as well as human-centered values and fairness, robustness, security and safety. The UNESCO recommendations on the ethics of AI follow this trend and enunciate the values and principles that should rule AI, including the protection of human rights, the environment and ecosystem, diversity and inclusion and the promotion of an interconnected just society. The principles are: proportionality and do no harm, fairness and non-discrimination, safety and security, sustainability, right to privacy and data protection, human oversight and determination, transparency and explainability, responsibility and ac-

countability, literacy, and multistakeholder collaboration. Similarly, the AI Bill of Rights, created by the United States White House Office of Science and Technology Policy, are guidelines for the responsible design and use of artificial intelligence. These guidelines include five principles: Safe and Effective Systems, Algorithmic Discrimination Protections, Data Privacy, Notice and Explanation, Human Alternatives, Consideration, and Fallback. Finally, the Hiroshima G7 AI process takes a similar approach developing a set of principles on AI governance, interoperability, and trustworthiness. Ongoing discussions could also include IP rights protections, transparency, and misinformation. G7 members work closely with the OECD and the GPAI.

The EU AI Act of 2021 (based on the AI whitepaper by the EU Commission), advances horizontal legislation to regulate AI from a precautionary approach based on risks. The rules were meant to ensure that AI develops and is used in Europe in line with EU rights and values including human oversight, safety, privacy, transparency, non-discrimination and social and environmental well being. The EU proposal identifies different types of risks as unacceptable risks, high risks and limited risks. It seeks to adopt the AI Act by the end of 2023. A similar approach seems to be proposed for Canada's AIDA which has yet to be drafted.

Separately, China has enacted Interim Administrative Measures for Generative Artificial Intelligence Services which entered into force in August 15 2023, and place a variety of responsibilities on generative AI service providers related to aspects such as AI governance, training data requirements, security, tagging and labeling standards, content moderation, data protection protocols, and safeguarding user rights related to privacy and intellectual property.

The proliferation of initiatives on AI shows increasing concerns over this technology and a perception of lack of existing standards to address them. Most of these proposals address AI as an industrial development. These industry-wide approaches may be useful as professional and industry standards mindful of the importance of care and caution as well as opportunities and risks that may arise from scientific breakthroughs. Still, these industry-wide approaches provide few guidelines for nuanced implementation of ex-

isting AI technologies. Rather, while concerned with industry ethics, they have important shortcomings to address and properly account for the opportunities and challenges that AI raises for specific human rights. To clarify, **while developing principles that guide AI development and progress is needed and important, it doesn't substitute for the need for nuanced, human rights based analysis of the concrete opportunities and challenges that AI implementation raises in different sectors.** While the principles governing the technology may be the same, the use of AI for self-driving cars is rather different than the use of AI for judicial decision making or the use of AI for private content moderation. Three very different uses and implications of the same technology.

Initiatives that focus on AI as a technological or scientific development tend to revolve around ethics and focus almost exclusively on risks. Adopting an industry-like approach may jeopardize the value of the human rights framework as the guiding framework in the development and deployment of these technologies.

We propose that the UN Secretary General and its Tech Envoy take a different approach considering its specific expertise and its potential impact on State policy worldwide. Human rights standards and practice offer valuable guidelines and insights for the creation of normative principles to rule the development of technology. International standards bind governments and corporations alike. Human rights offer a shared discourse and practice that has decades in the making and the UN has invested heavily on its development and domestic adoption to this day.

While there may be other organizations better positioned to develop ethical guidelines for the industry of AI, the UN is uniquely positioned to develop concrete, human rights oriented standards for the development, regulation and deployment of AI. The complex architecture of UN treaty bodies, global human rights experts, the Office of the High Commissioner for Human Rights, provides the UN with a significant advantage and a particularly solid trajectory to guide both the development and the regulation of these new technologies per its historically held standards.

This submission poses that any UN governing attempt or mechanism of AI should squarely build on the existing UN human rights framework, tak-

ing into account how different rights are enforced and how they are doctrinally and jurisprudentially developed to dialogue with one another —often in tensions that pose serious challenges but that are ultimately creative.

A strategic approach to AI governance: Granular approach of AI from a human rights perspective

Rather than an ethical framework to lead the industry or a risk-based approach to technology, the UN should promote a legal framework, more granular and rights specific. If positive and negative impacts of AI are to be contemplated for analysis, the initiative should be human rights centered rather than risk-centered.

AI poses potential risks and opportunities for human development. Addressing AI from a rights-based perspective would contribute to better identifying the potential positive and negative implications of AI for society. A common problem with industry-wide approaches to AI is that they assume there is a single source of potential risk and a single target to find the solution. We contend that AI poses different challenges according to the field where it is deployed. For instance, the impacts that AI may have on the right to freedom of expression are not the same as the impacts it may have on the right of civilians to be protected from unmanned attacks under international humanitarian law. Even if we could agree on a horizontal transparency mandate, those obligations would be quite different for autonomous weapon systems than those of a large language model-based chatbot.

AI, as (un)defined today, encompasses many things that, though connected, are hardly the same. And different things are treated differently in law, politics, economics, and industry. The risks-based approach often-times ignores those differences and nuances. Creating a fixed hierarchy of risks (i.e., some that are admissible and others that are unacceptable) independent of where or how AI is used or implemented can create distortions and abuse, leading to under and over protected areas in different fields with significant impact to human rights.

Furthermore, a risk-based approach frames the issue in a problem-based

perspective (e.g., disinformation). Such an approach requires consensus on the diagnostics and the existing problems and common elements to those problems across the spectrum, which today, we argue, are lacking. There are important gaps in our evidence and understanding of how technology **impacts** human behavior and there are important differences in the understanding of technology generated risks across borders (particularly across regions). The risk approach fails to acknowledge human rights as the framework but rather factors in human rights at the end stage or can even lead to results at odds with human rights. For instance, disinformation is an issue that has been linked to AI and is a concern in this area. Some proposals are suggesting adding watermarks or embedding metadata indicating the authorship of AI-generated content as an effective way of fighting deep fakes and misinformation. While the solution can be viable and ethical from a technological and industry perspective, it fails to question whether it is proportional or even necessary and fails to acknowledge the impact such a measure would have on anonymity or privacy for that matter. It is usually those who are closer to the issues of concern who can better identify the new challenges and the trade offs that AI and its regulation supposes.

Process: Creation of a global forum/dialogue with different tracks under the auspices of the United Nations

The substantive vision presented in the previous section could be put into practice by creating mechanisms for coordination and dialogue among already existing UN human rights fora. The UN human rights infrastructure (i.e. UNESCO, the system of Rapporteurs, thematic working groups, etc) could and should be capitalized for this work as well. Doing so would even guarantee certain continuity to the process, cohesiveness to the products generated and provide the structure to update or follow up when required. Moreover, it would prevent the voices of such global UN appointed thematic experts from being silenced or cut entirely in faceless, referendum-like consultations where quantity can overturn and overshadow quality towards results that are at odds with international human rights standards.

The United Nations as a global organization is uniquely positioned to lead

a global dialogue and summon its own international human rights experts to discuss and analyze the implications of AI for human rights and society.

If a new forum is to be created, we believe it should be developed in tracks. Past and comparative experiences could help shape such effort. Structures like the G20 or ICANN can contribute best and worst practices as well as lessons learnt even if they serve different purposes. One interesting aspect is that they both propose a tracks-based approach. The G20 model is organized in tracks which are then divided into more specific working groups within each track. ICANN, for its part, has a structure of topic-based Support Organizations that develop policies, that are in turn subject to the recommendations of Advisory Committees that represent [multiple stakeholders and interests](#). ICANN also has a public consultation process for decision-making that [ensures the participation of all interested or affected groups within its structure](#) (including an Advisory Committee that represents end users).

Conclusion

There is increasing concern over the challenges and risks potentially posed by AI on several values and rights. So far there has been a heavy trend to try to address these potential challenges by regulating or attempting to frame the technology itself. We believe that general ethical principles must guide the development of AI. However, new efforts by the UN should capitalize on existing expertise and strengths and provide much needed guidelines and standards for the use and deployment of such technology from a human rights perspective. In order to do so, **we believe and recommend that the UN take a human rights centered rather than the human centered approach that based on existing documentation and discussion seems to be heavily focused on risks rather than rights.**

If a new process for IA is adopted, tracks could be constructed around specific rights and should be led by UN experts. Their voices should be heard and valued in designing the new forum and they should have a hierarchical voice to lead and frame the conversation and identify the red lines that could prevent recommendations and conclusions to be adopted. A struc-

ture like this would also guarantee that all basic, interrelated human rights standards are incorporated into the design itself. Each track could have specific working groups that could dialogue with each other. For example, there could be a track on freedom of expression and AI and a working group, within this track, on content moderation and AI. This model could and should also allow for the participation of other relevant actors such as States, civil society, academia and the private sector and there are already best and worst practices to guarantee meaningful participation within the UN as well as elsewhere that should be taken into account.

For more information, you may contact us at cele@palermo.edu.