16 July 2024

UNGA adopts first resolution on Artificial Intelligence

By Viviana Munoz Tellez

The United Nations recently approved a first resolution on Artificial Intelligence (AI). It contains a number of important principles and objectives that if achieved can help to leverage the potential of AI systems in all countries and control their risks. However, issues of critical importance for developing countries, such as bridging the digital divide in the use of AI, capacity building, ethics, bias and unfair data exploitation, are not adequately covered.

Les Nations Unies ont récemment approuvé une première résolution sur l'intelligence artificielle (IA). Elle contient plusieurs principes et objectifs importants qui, une fois atteints, pourront contribuer à exploiter le potentiel des systèmes d'IA dans tous les pays et à contrôler les risques associés à ces systèmes. Toutefois, des questions d'une importance cruciale pour les pays en développement, telles que la réduction de la fracture numérique dans l'utilisation de l'IA, le renforcement des capacités, l'éthique, les préjugés et l'exploitation déloyale des données, ne sont pas abordées de manière adéquate.

Las Naciones Unidas han aprobado recientemente una primera resolución sobre Inteligencia Artificial (IA). Contiene una serie de principios y objetivos importantes que, de alcanzarse, pueden ayudar a aprovechar el potencial de los sistemas de IA en todos los países y a controlar sus riesgos. Sin embargo, no se abordan adecuadamente cuestiones de vital importancia para los países en desarrollo, como la reducción de la brecha digital en el uso de la IA, la creación de capacidades, la ética, la parcialidad y la explotación injusta de los datos.

SOUTH CENTRE

There is significant optimism about the development of artificial intelligence (AI). However, concern about the ability of AI to do wrong, or even to replace the human species or annihilate it, are growing. AI is a very powerful technology. We are training AI systems and then learning about what capabilities they are developing. Autonomous AI systems already exist. More resources are going into training AI than to ensuring safe AI. The extent to which humans will be able to control how it evolves, as it becomes smarter, and the extent of possible misuse of AI, are big open questions. General super intelligent AI systems may become uncontrollable. Regulation of AI is imperative.

The United Nations (UN) General Assembly recently adopted the first resolution on artificial intelligence, A/78/L.49. The resolution focuses on the role that the United Nations system can play to reach global consensus on safe, secure and trustworthy AI systems. These should be consistent with international law, in particular the Charter of the United Nations; the Universal Declaration of Human Rights; and the 2030 Agenda for Sustainable Development, including by promoting inclusive international cooperation and facilitating the inclusion, participation and representation of developing countries in deliberations.

The resolution establishes some general principles for AI systems to do good, while recognising some of the risks. The resolution highlights the need to address AI and digital divides, improve the governance of AI, strengthen capacities and increase financing to developing countries. However, the way forward for these transformations is not proposed. In the area of regulating AI towards ensuring safety, the resolution advances some limited recommendations.

The resolution notes the relevance of cooperation in research and development of AI both to advance its benefits and address its risks, but it misses to highlight the importance of fostering open research and open source as the best ways to mitigate the risks of AI, by promoting testing by broad communities and debugging of AI. The resolution does not differentiate between types of AI, in particular narrow AI tools as opposed to super smart AI systems that can have agency, i.e., consciousness and self-awareness that can have the ability to take control. The risks of these AI systems are difficult to even conceive.

The resolution frames the development of AI in a positive manner, as "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development", but some of the risks are also recognised, though not sufficiently.

The resolution recognises that safe, secure and trustworthy AI systems, (not including AI systems in the military domain) have the potential to accelerate and enable progress towards the achievement of all the Sustainable Development Goals (SDGs) and sustainable development, promote digital transformation, promote peace, overcome digital divides between and within countries, and promote and protect the enjoyment of human rights. The resolution also recognizes that the improper or malicious design, development, deployment and use of AI systems, such as without adequate safeguards or in a manner inconsistent with international law, pose risks that could hinder progress towards the achievement of the 2030 Agenda for Sustainable Development and the SDGs and undermine sustainable development, widen digital divides between and within countries, reinforce structural inequalities and biases, lead to discrimination, undermine information integrity and access to information, undercut the protection, promotion and enjoyment of human rights and fundamental freedoms, including the right not to be subject to unlawful or arbitrary interference with one's privacy, and increase the potential risk for accidents and compound threats from malicious actors.

The resolution also stresses the urgency of achieving global consensus on safe, secure and trustworthy AI systems, facilitating inclusive international cooperation to formulate and use effective, internationally interoperable safeguards, practices and standards that promote innovation and prevent the fragmentation of the governance of safe, secure and trustworthy AI systems.

The existing AI and other digital divides, and the varying levels of technological development between and within countries are noted, as well as the unique challenges that developing countries face in keeping pace with this rapid acceleration and the need to narrow the existing disparities between developed and developing countries in terms of conditions, possibilities and capacities, and the urgency to strengthen capacity building and technical and financial assistance to developing countries to close digital divides between and within countries. The resolution thus underlines the importance of bridging the AI and other digital divides between and within countries.

General principles for data governance are advanced. The resolution emphasizes on the fair, inclusive, responsible and effective data governance, improving data generation, accessibility and infrastructure, and the use of digital public goods. Promote international cooperation, collaboration and assistance on data governance for greater consistency and interoperability, where feasible, of approaches for advancing trusted cross-border data flows for safe, secure and trustworthy AI systems, and make its development more inclusive, equitable, effective and beneficial to all.

The governance of AI is also addressed. The need for effective, equitable and meaningful participation and representation of developing countries in international processes and forums on the governance of AI systems, is recognised. Possible AI governance approaches are still in discussion. The resolution provides general principles only, as appropriate, based on international law, interoperable, agile, adaptable, inclusive, responsive to the different needs and capacities of developed and developing countries alike and for the benefit of all, as the technology and our understanding of it develops.

The resolution resolves to promote safe, secure and trustworthy AI systems to accelerate progress towards the full realization of the 2030 Agenda for Sustainable Development, stressing the need for a standard of safe, secure and trustworthy AI systems that promote, not hinder, digital transformation and equitable access to their benefits in order to achieve the set goals. It does not elaborate on how this will be promoted.

The resolution addresses Member States as well as multi-stakeholders to work in partnership, including from the private sector, international and regional organizations, civil society, the media, academia and research institutions, and technical communities and individuals, to develop and support regulatory and governance approaches and frameworks related to safe, secure and trustworthy AI systems.

Member States and stakeholders are asked to cooperate with and provide assistance to developing countries towards inclusive and equitable access to the benefits of digital transformation and safe, secure and trustworthy AI systems, including by: (a) expanding participation in digital transformation, including by capacity building relating to AI systems, promoting knowledge sharing activities and the transfer of technology on mutually agreed terms, (b) enhancing digital infrastructure connectivity and access to technological innovations, (c) enhancing the ability of developing countries, in particular the least developed countries, to address major structural impediments and lift obstacles to accessing the benefits of new and emerging technologies and AI innovation, (d) aiming to increase funding for SDGs related research and innovation related to digital technologies and safe, secure and trustworthy AI systems and build capacity in all regions and countries to contribute to and benefit from this research, (e) enabling international innovation-based environments to enhance the ability of developing countries to develop technical expertise and capacities, harness

data and computing resources, and national regulatory and governance approaches, frameworks and procurement capacity, and creating an inclusive enabling environment at all levels for safe, secure and trustworthy AI systems-based solutions, (f) urgently mobilizing means of implementation such as technology transfer on mutually agreed terms, capacity building to close the AI and other digital divides, technical assistance and financing to developing countries related to AI systems, (g) promoting the access to and design, development, deployment, and use of safe, secure and trustworthy AI systems.

Concerning the regulation of AI, the resolution advances a few elements. It calls upon Member States and stakeholders to refrain from or cease the use of AI systems that are impossible to operate in compliance with international human rights law or that pose undue risks to the enjoyment of human rights and reaffirms that the same rights that people have offline must also be protected online, including throughout the life cycle of AI systems. It also encourages all Member States and other stakeholders to promote the development and implementation of domestic regulatory and governance approaches and frameworks to support responsible and inclusive AI innovation and investment for sustainable development, while simultaneously promoting safe, secure and trustworthy AI systems. Moreover, Member States and stakeholders are to encourage effective measures that promote innovation for the internationally interoperable identification, classification, evaluation, testing, prevention and mitigation of vulnerabilities and risks during the design and development and prior to the deployment and use of AI systems, and the incorporation of feedback mechanisms to allow evidence based discovery and reporting by end-users and third parties of technical vulnerabilities and, as appropriate, misuses of AI systems and AI incidents following their development, testing and deployment to address them. It also calls on Member States and stakeholders to foster the development, implementation and disclosure of mechanisms of risk monitoring and management, mechanisms for securing data, including personal data protection and privacy policies, as well as impact assessments as appropriate, across the life cycle of AI systems, and strengthening investment in developing and implementing effective safeguards, including physical security, AI systems security, and risk management across the life cycle of AI systems.

The resolution also encourages the development and deployment of effective, accessible, adaptable, internationally interoperable technical tools, standards or practices, including reliable content authentication and provenance mechanisms – such as watermarking or labelling, where technically feasible and appropriate, that enable users to identify information manipulation, distinguish or determine the origins of authentic digital content and AI-generated or manipulated digital content – and increasing media and information literacy.

The resolution further calls to facilitate the development and implementation of effective, internationally interoperable frameworks, practices and standards for training and testing AI systems to enhance policymaking and to help protect individuals from all forms of discrimination, bias, misuse or other harm, and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of AI systems, including, for example, by analysing and mitigating bias encoded in datasets and otherwise combating algorithmic discrimination and bias, while not inadvertently or disproportionally impacting the positive development, access and uses of other users and beneficiaries. It calls for promoting transparency, predictability, reliability and understandability throughout the life cycle of AI systems that make or support decisions impacting

end-users, including providing notice and explanation, and promoting human oversight, such as, for example, through review of automated decisions and related processes or, where appropriate and relevant, human decision – making alternatives or effective redress and accountability for those adversely impacted by automated decisions of artificial intelligence systems.

Importantly, the resolution also calls for strengthening investment in developing and implementing effective safeguards, including risk and impact assessments, throughout the life cycle of AI systems to protect the exercise of and mitigate against the potential impact on the full and effective enjoyment of human rights and fundamental freedoms.

The resolution encourages the private sector to adhere to applicable international and domestic laws and act in line with the United Nations Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, recognizes the need for increased collaboration, including between and within the public and private sectors and civil society, academia and research institutions and technical communities, to provide and promote fair, open, inclusive and non-discriminatory business environment, economic and commercial activities, competitive ecosystems and marketplaces across the life cycle of safe, secure and trustworthy AI. The resolution also encourages Member States to develop policies and regulations to promote competition in safe, secure and trustworthy AI systems and related technologies.

The resolution is timely in the context of the current negotiations for a Global Digital Compact and the UN Summit of the Future to be held in September 2024. The resolution contains a number of important principles and objectives that if achieved can help to leverage the potential of AI systems in all countries and control their risks. The adoption of AI guidelines and regulations in countries such as the United States of America, the European Union and China opens questions about the likelihood of finding common approaches in line with the recommendations of the UN resolution (e.g., regarding interoperability and data privacy). The negotiations of the Global Digital Compact offer an opportunity to move from principles and objectives to more concrete outcomes concerning **how** those principles can be implemented, and the objectives achieved, including how to deal with the technological and economic power of the major digital companies and prevent anti-competitive practices.

Author: Viviana Munoz Tellez is Coordinator of the Health, Intellectual Property and Biodiversity Programme (HIPB) of the South Centre.

SOUTHVIEWS is a service of the South Centre providing opinions and analysis of topical issues from a South perspective.

The views contained in this article are attributable to the author(s) and do not represent the institutional views of the South Centre or its Member States.

For more information, please contact Anna Bernardo of the South Centre:
Email abernardo@southcentre.int, or telephone +41 22 791 8050.

Front cover photo credit: Ake Puttisarn from Getty Images