General Assembly

Seventy-eighth session
Agenda item 13
**Integrated and coordinated implementation of and follow-up
to the outcomes of the major United Nations conferences and
summits in the economic, social and related fields**

Resolution adopted by the General Assembly on 21 March 2024

[*without reference to a Main Committee (A/78/L.49)*]

78/265. Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development

The General Assembly,

Reaffirming international law, in particular the Charter of the United Nations, and recalling the Universal Declaration of Human Rights,[1]

Reaffirming also its resolution 70/1 of 25 September 2015, entitled "Transforming our world: the 2030 Agenda for Sustainable Development", its resolution 69/313 of 27 July 2015 on the Addis Ababa Action Agenda of the Third International Conference on Financing for Development, and the political declaration adopted at the high-level political forum on sustainable development convened under the auspices of the General Assembly contained in the annex to its resolution 78/1 of 29 September 2023,

Recalling its resolutions 77/320 of 25 July 2023 on the impact of rapid technological change on the achievement of the Sustainable Development Goals and targets, 78/132 of 19 December 2023 on information and communications technologies for sustainable development, 78/160 of 19 December 2023 on science, technology and innovation for sustainable development, 78/213 of 19 December 2023 on the promotion and protection of human rights in the context of digital technologies, 77/211 of 15 December 2022 on the right to privacy in the digital age, 70/125 of 16 December 2015 on the overall review of the implementation of the outcomes of the World Summit on the Information Society, all the outcomes of the World Summit on the Information Society, including the Geneva Declaration of Principles,[2] the Geneva Plan of Action,[3]

_____

[1] Resolution 217 A (III).
[2] See A/C.2/59/3, annex.
[3] Ibid.

A/RES/78/265

Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development

the Tunis Commitment[4] and the Tunis Agenda for the Information Society,[5] and the declaration on the commemoration of the seventy-fifth anniversary of the United Nations contained in its resolution 75/1 of 21 September 2020,

Taking note of the efforts of the International Telecommunication Union, in partnership with 40 United Nations bodies, to convene the Artificial Intelligence for Good platform, including its annual summit and the launch of the International Telecommunication Union's Artificial Intelligence Repository to identify responsible and practical applications of artificial intelligence to advance the Sustainable Development Goals; and the adoption by the General Conference of the United Nations Educational, Scientific and Cultural Organization of its Recommendation on the Ethics of Artificial Intelligence of 23 November 2021,[6] its implementation plan, including the Readiness Assessment Methodology and the Ethical Impact Assessment and the Global Forum on the Ethics of Artificial Intelligence; as well as taking note of the Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework,[7] as endorsed by the Human Rights Council in its resolution 17/4 of 16 June 2011;[8] and the work of the Office of the United Nations High Commissioner for Human Rights regarding artificial intelligence,

Taking note also of the report of the Secretary-General of the United Nations entitled "Road map for digital cooperation",[9] as well as the establishment of the Office of the Secretary-General's Envoy on Technology to coordinate its implementation and the establishment by the Secretary-General of a multi-stakeholder High-level Advisory Body on Artificial Intelligence and of its interim report issued on 21 December 2023, and looking forward to its final report,

Recognizing that safe, secure and trustworthy artificial intelligence systems – which, for the purpose of this resolution, refers to artificial intelligence systems in the non-military domain, whose life cycle includes the stages: pre-design, design, development, evaluation, testing, deployment, use, sale, procurement, operation and decommissioning, are such that they are human-centric, reliable, explainable, ethical, inclusive, in full respect, promotion and protection of human rights and international law, privacy preserving, sustainable development oriented, and responsible – have the potential to accelerate and enable progress towards the achievement of all 17 Sustainable Development Goals and sustainable development in its three dimensions – economic, social and environmental – in a balanced and integrated manner; promote digital transformation; promote peace; overcome digital divides between and within countries; and promote and protect the enjoyment of human rights and fundamental freedoms for all, while keeping the human person at the centre,

Recognizing also that the improper or malicious design, development, deployment and use of artificial intelligence 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 its Sustainable Development Goals and undermine sustainable development in its three dimensions – economic, social and environmental; 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

---

[4] See A/60/687.

[5] Ibid.

[6] United Nations Educational, Scientific and Cultural Organization, *Records of the General Conference, Forty-first Session, Paris, 9–24 November 2021*, vol. 1, *Resolutions*, annex VII.

[7] A/HRC/17/31, annex.

[8] See *Official Records of the General Assembly, Sixty-sixth Session, Supplement No. 53* (A/66/53), chap. III, sect. A.

[9] A/74/821.

Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development

A/RES/78/265

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,

Recognizing further the rapid acceleration of the design, development, deployment and use of artificial intelligence systems and rapid technological change, and their potential impact in accelerating the achievement of the Sustainable Development Goals, therefore stressing the urgency of achieving global consensus on safe, secure and trustworthy artificial intelligence 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 artificial intelligence systems; and recognizing also existing artificial intelligence and other digital divides, and the varying levels of technological development between and within countries, that developing countries face unique challenges in keeping pace with this rapid acceleration, which cause obstacles to sustainable development, the need to narrow the existing disparities between developed and developing countries in terms of conditions, possibilities and capacities, therefore stressing also the urgency of strengthening capacity building and technical and financial assistance to developing countries to close digital divides between and within countries and support developing countries' effective, equitable and meaningful participation and representation in international processes and forums on the governance of artificial intelligence systems,

Recognizing also that the governance of artificial intelligence systems is an evolving area and the need for continued discussions on possible governance approaches that are 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,

1.   *Resolves* to bridge the artificial intelligence and other digital divides between and within countries;

2.   *Resolves* to promote safe, secure and trustworthy artificial intelligence systems to accelerate progress towards the full realization of the 2030 Agenda for Sustainable Development, [10] further bridging the artificial intelligence and other digital divides between and within countries; and stresses the need for the standard of safe, secure and trustworthy artificial intelligence systems to promote, not hinder, digital transformation and equitable access to their benefits in order to achieve all 17 Sustainable Development Goals and sustainable development in its three dimensions – economic, social and environmental – and address other shared global challenges, particularly for developing countries;

3.   *Encourages* Member States and invites multi-stakeholders from all regions and countries, within their respective roles and responsibilities, 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 artificial intelligence systems that create an enabling ecosystem at all levels, including for innovation, entrepreneurship and the dissemination of knowledge and technologies on mutually agreed terms, recognizing that effective partnership and cooperation between Governments and multi-stakeholders is necessary in developing such approaches and frameworks;

4.   *Calls upon* Member States and invites other stakeholders to take action to cooperate with and provide assistance to developing countries towards inclusive and

_____
[10] Resolution 70/1.

equitable access to the benefits of digital transformation and safe, secure and trustworthy artificial intelligence systems, including by:

(a) Expanding participation of all countries, in particular developing countries, in digital transformation to harness the benefits and effectively participate in the development, deployment and use of safe, secure and trustworthy artificial intelligence systems, including by capacity building relating to artificial intelligence systems, recognizing that promoting knowledge sharing activities and the transfer of technology on mutually agreed terms is an important aspect of building capacity, stressing the need to close the artificial intelligence and other digital divides; and increase digital literacy;

(b) Enhancing digital infrastructure connectivity and access to technological innovations through stronger partnerships to help developing countries effectively participate throughout the life cycle of artificial intelligence systems and accelerate the inclusive and positive contribution of artificial intelligence systems to society, including towards the full realization of the 2030 Agenda and its Sustainable Development Goals, while ensuring that artificial intelligence systems around the world are safe, secure and trustworthy throughout their life cycle;

(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 artificial intelligence innovation to achieve all 17 Sustainable Development Goals, including through scaling up the use of scientific sources, affordable technology, research and development, including through strengthened partnerships;

(d) Aiming to increase funding for Sustainable Development Goals related research and innovation related to digital technologies and safe, secure and trustworthy artificial intelligence 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 compute resources, and national regulatory and governance approaches, frameworks and procurement capacity, and create an inclusive enabling environment at all levels for safe, secure and trustworthy artificial intelligence systems-based solutions;

(f) Urgently mobilizing means of implementation such as technology transfer on mutually agreed terms, capacity building to close the artificial intelligence and other digital divides, technical assistance and financing to developing countries related to artificial intelligence systems in accordance with developing countries' national needs, policies and priorities;

(g) Promoting the access to and design, development, deployment and use of safe, secure and trustworthy artificial intelligence systems to achieve sustainable development in its three dimensions – economic, social and environmental;

5. *Emphasizes* that human rights and fundamental freedoms must be respected, protected and promoted throughout the life cycle of artificial intelligence systems, calls upon all Member States and, where applicable, other stakeholders to refrain from or cease the use of artificial intelligence systems that are impossible to operate in compliance with international human rights law or that pose undue risks to the enjoyment of human rights, especially of those who are in vulnerable situations, and reaffirms that the same rights that people have offline must also be protected online, including throughout the life cycle of artificial intelligence systems;

Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development

A/RES/78/265

6.   *Encourages* all Member States, where appropriate, in line with their national priorities and circumstances and while implementing their distinct national regulatory and governance approaches and frameworks, and, where applicable, other stakeholders to promote safe, secure and trustworthy artificial intelligence systems in an inclusive and equitable manner, and for the benefit of all, and foster an enabling environment for such systems to address the world's greatest challenges, including achieving sustainable development in its three dimensions – economic, social and environmental – with specific consideration of developing countries and leaving no one behind by:

(a)   Promoting the development and implementation of domestic regulatory and governance approaches and frameworks, in line with their respective national, and where applicable subnational, policies and priorities and obligations under international law, to support responsible and inclusive artificial intelligence innovation and investment for sustainable development, while simultaneously promoting safe, secure and trustworthy artificial intelligence systems;

(b)   Encouraging 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 artificial intelligence systems;

(c)   Encouraging 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 artificial intelligence systems and artificial intelligence incidents following their development, testing and deployment to address them;

(d)   Raising public awareness and understanding of the core functions, capabilities, limitations and domains of appropriate civil use of artificial intelligence systems;

(e)   Fostering 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 artificial intelligence systems;

(f)   Strengthening investment in developing and implementing effective safeguards, including physical security, artificial intelligence systems security, and risk management across the life cycle of artificial intelligence systems;

(g)   Encouraging 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 artificial intelligence-generated or manipulated digital content – and increasing media and information literacy;

(h)   Facilitating the development and implementation of effective, internationally interoperable frameworks, practices and standards for training and testing artificial intelligence 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 artificial intelligence 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;

A/RES/78/265

Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development

(i)   Encouraging, where appropriate and relevant, the implementation of appropriate safeguards to respect intellectual property rights, including copyright-protected content, while promoting innovation;

(j)   Safeguarding privacy and the protection of personal data when testing and evaluating systems, and for transparency and reporting requirements in compliance with applicable international, national and subnational legal frameworks, including on the use of personal data throughout the life cycle of artificial intelligence systems;

(k)   Promoting transparency, predictability, reliability and understandability throughout the life cycle of artificial intelligence 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;

(l)   Strengthening investment in developing and implementing effective safeguards, including risk and impact assessments, throughout the life cycle of artificial intelligence systems to protect the exercise of and mitigate against the potential impact on the full and effective enjoyment of human rights and fundamental freedoms;

(m)   Promoting artificial intelligence systems that advance, protect and preserve linguistic and cultural diversity, taking into account multilingualism in their training data and throughout the life cycle of the artificial intelligence system, particularly for the large language models;

(n)   Intensifying information-sharing on mutually agreed terms among entities with roles across the life cycle of artificial intelligence systems to identify, understand and act using scientific and evidence-based best practices, policies and approaches to artificial intelligence systems to maximize the benefits while mitigating the potential risks across the life cycle of artificial intelligence systems, including advanced artificial intelligence systems;

(o)   Encouraging research and international cooperation to understand, balance and address the potential benefits and risks related to the role of artificial intelligence systems in bridging digital divides and achieving all 17 Sustainable Development Goals, including the role of scaling up of digital solutions such as open-source artificial intelligence systems;

(p)   Calling upon Member States to adopt specific measures to close the gender digital divide and to ensure that particular attention is paid to access, affordability, digital literacy, privacy and online safety, to enhance the use of digital technologies, including artificial intelligence systems, and to mainstream a disability, gender and racial equality perspective in policy decisions and the frameworks that guide them;

(q)   Encouraging research and international cooperation to develop measures for the identification and assessment of the impacts of the deployment of artificial intelligence systems on labour markets, and providing support for the mitigation of potential negative consequences for workforces, especially in developing countries, in particular the least developed countries, and fostering programmes aimed at digital training, capacity building, supporting innovation and enhancing access to benefits of artificial intelligence systems;

7.   *Recognizes also* that data is fundamental to the development and operation of artificial intelligence systems; emphasizes that the fair, inclusive, responsible and effective data governance, improving data generation, accessibility and infrastructure, and the use of digital public goods are essential to harnessing the potential of safe,

Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development

A/RES/78/265

secure and trustworthy artificial intelligence systems for sustainable development, and urges Member States to share best practices on data governance and to 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 artificial intelligence systems, and make its development more inclusive, equitable, effective and beneficial to all;

8. *Acknowledges* the importance of continuing the discussion on developments in the area of artificial intelligence governance so that international approaches keep pace with the evolution of artificial intelligence systems and their uses; and encourages continued efforts by the international community to promote inclusive research, mapping and analysis that benefit all parties on the potential impacts and applications that artificial intelligence systems and rapid technological change can have in the development of existing and new and emerging technologies and on accelerating the achievement of all 17 Sustainable Development Goals, and to inform how to develop, promote and implement effective, internationally interoperable safeguards, practices, standards and tools for artificial intelligence designers, developers, evaluators, deployers, users and other stakeholders for safe, secure and trustworthy artificial intelligence systems; as well as stresses the need for Governments, the private sector, civil society, international and regional organizations, academia and research institutions and technical communities and all other stakeholders to continue to work together, as appropriate; as well as acknowledges the need for more cohesive, effective, coordinated and inclusive engagement and participation of all communities, particularly from developing countries, in the inclusive governance of safe, secure and trustworthy artificial intelligence systems;

9. *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; acknowledges the importance of more inclusive and equitable access to the benefits of safe, secure and trustworthy artificial intelligence systems; and 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 artificial intelligence; as well as encourages Member States to develop policies and regulations to promote competition in safe, secure and trustworthy artificial intelligence systems and related technologies, including by supporting and enabling new opportunities for small businesses and entrepreneurs and technical talent, and enabling fair competition in the artificial intelligence marketplace, through critical investment, especially for developing countries;

10. *Calls upon* specialized agencies, funds, programmes, other entities, bodies and offices, and related organizations of the United Nations system, within their respective mandates and resources, to continue to assess and enhance their response to leverage the opportunities and address the challenges posed by artificial intelligence systems in a collaborative, coordinated and inclusive manner, through appropriate inter-agency mechanisms, including by conducting research, mapping and analysis that benefit all parties on the potential impacts and applications; reporting on progress and challenges in addressing issues; and cooperating with and assisting developing countries in capacity building, access and sharing the benefits of safe, secure and trustworthy artificial intelligence systems in achieving all 17 Sustainable Development Goals and sustainable development in its three dimensions – economic, social and environmental; stressing the need to close artificial intelligence and other digital divides between and within countries;

A/RES/78/265

Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development

11.   *Recalls* its resolution 76/307 of 8 September 2022 and its decision 77/568 of 1 September 2023 on the modalities and the scope of the Summit of the Future and, in this regard, looks forward to the development of a global digital compact;

12.   *Looks forward also* to the overall review by the General Assembly, in 2025, of the progress made since the World Summit on the Information Society;

13.   *Acknowledges* that the United Nations system, consistent with its mandate, uniquely contributes to reaching global consensus on safe, secure and trustworthy artificial intelligence systems, that is 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.

63rd plenary meeting
21 March 2024

———————