

COMMITTEE: Human Rights II

QUESTION OF: State censorship and surveillance in the age of artificial intelligence

SUBMITTED BY: Brazil

CO-SUBMITTED BY: USA, Israel, Finland, Italy

SIGNATORIES: Canada, France, Bangladesh, South Korea, Turkey, Sierra Leone, Azerbaijan, Ghana, UK

The General Assembly,

Recalling the International Covenant on Civil and Political Rights (ICCPR) Article 17 and Article 19, and General Comment No. 34 clarifying the core legal guidelines that we must acknowledge and ensure compliance with when contemplating any state limitation on expression or surveillance,

Reaffirming international standards on freedom of expression and privacy articulated by UNESCO, including the three-part test that governs restrictions under Article 19(3),

Recognizing the OECD AI Principles that call for transparency, explainability, human-rights respect, accountability and oversight across the AI lifecycle, offering an interoperable baseline for trustworthy AI use across sectors,

Concerned by misuse of AI-enabled surveillance and the risk to journalists, human rights defenders and civilians, as emphasized by UN special procedure experts,

Alarmed that unregulated deployment of biometric analytics, real-time facial recognition, predictive policing and bulk data collection can produce discrimination and unlawful interference with privacy absent robust safeguards, as underscored by OHCHR's "privacy in the digital age" work,

Mindful of current domestic debates, including the pending AI regulation bill advancing a risk-based approach akin to the EU AI Act, focused on transparency and responsibility in moderation in order to comply with international law,

1. Encourages all member states to adopt and/or update a digital rights baseline in order to achieve necessary freedoms and protection,
 - a. These documents should be actively codifying online freedoms and privacy in an internet rights framework;
 - i. Focusing on freedom of expression, net neutrality, privacy, personal security, due process, multistakeholder governance, and other issues specific to each member states struggles and trends,
 - b. Guaranteeing due process for content restrictions, including notice, reasoned decisions, and access to appeal, to avoid covert state censorship in accordance to the UN and OECD norms with reducing arbitrary censorship;
 - c. Affirms that the establishment of safety constraints on AI systems does not equate to censorship where such restrictions protect against misinformation, violence, discrimination, or violations of fundamental rights;
2. Encourages the creation of a framework for AI public safety & rights to govern AI-enabled surveillance and content governance:
 - a. Codifying AI surveillance and content-restriction powers in publicly accessible law with clear definitions, limited purposes, algorithmic takedowns (public safety, serious crime), and time-bound mandates;
 - i. The definitions will majorly involve tackling topics like facial recognition,
 - ii. Limited purposes refers mainly to predictive policing,
 - iii. Algorithm takedowns will focus on only serious crimes and violations of national law and/or the above mentioned Digital Rights Baseline for the strict purpose of public safety as to avoid any exploitation,
 - b. Insuring independent authorization through requiring ex ante authorization by a competent independent authority for high-risk deployments;
 - i. Sourcing this independent authority on a local level and involving data protection or judicial body,
 - c. Establishing minimum safeguards within this documentation including various methods and processes,
 - i. Including the use of necessity-proportionality assessments, strict data minimization, access controls, and audit trails,

- ii. Additionally constantly looking into and staying updated on least-intrusive means which seem to majorly involve those mentioned above;
 - d. Constantly making use of human oversight through the use and creation of documented roles for accountable officials;
 - i. Requiring a human review panel whenever digital censorship is concerned,
 - 1. Providing resources and guidance if needed to areas that struggle with implementing this requirement in a meaningful form,
 - 2. Supplementing these with technical workshops at the Global Dialogue on AI Governance
 - e. Making use of the OHCHR reporting method to document violations of international law involving AI-enabled censorship, political manipulation, or violations of digital rights;
 - f. Solemnly affirms a multi-layered, rules-based system to combat state censorship and surveillance in the age of artificial intelligence that promotes:
 - i. Human ethics, Robust legal frameworks like the EU AI Act, International transparency and accountability, Comprehensive digital literacy education;
3. Suggests states to establish comprehensive data protection regimes with internal structures to oversee AI surveillance and content governance,
- a. Through this movement, additionally creating protections for the use of and storing of sensitive biometric data,
 - b. Ensuring mandatory or on-request DPIAs for high-risk processing and public guidance on DPIAs and high-risk criteria, following ANPD practice
 - i. Following such a structure as the ANPD demonstrated practical oversight mechanisms and DPIAs operationalize necessity and proportionality before deployment;
4. Requires pre-deployment Algorithmic/Data Protection Impact Assessments (AIA/DPIA) for any AI system affecting privacy, expression, assembly, or civic participation:
- a. Requiring AIA/DPIA before procurement or pilot; update on significant changes;
 - b. Making use of systematic description, risk map, mitigation plan, metrics, and community;
 - i. Specifically ensuring consultation with minorities and journalist due to they way they are significantly targeted by this technology;

- c. Publishing public non-confidential summaries and providing the full reports directly to oversight bodies;
 - d. Creating mandatory triggers where biometrics, mass monitoring, automated decisions with significant effects, or large-scale processing (>2M subjects) occur;
 - e. ANPD and peer authorities issue a shared AIA/DPIA template and risk taxonomy
 - i. Insuring adoption within 9 months
 - f. Promoting a Risk-Based Global AI Governance Compact
 - i. Ensuring alignment with the OECD AI Principles
 - ii. Prohibiting excessive-risk AI uses unless exceptional, strictly time-bound, independently authorized situations appear,
 - 1. Excessive-risk AI mainly incorporating indiscriminate real-time facial recognition of the general public as well as emotion recognition for policing,
 - iii. High-risk AI will additionally require DPIA/AIA, human oversight, explainability, incident reporting, and independent audits;
5. Invites better balance innovation with proper safeguarding through regulatory systems for sandboxes and testing to better ensure safety from the start,
- a. Regarding sandboxing it is key to make use of oversight for limited-scope pilots with summaries released publicly and well-detailed to inform the global community and ensure both ethical use and creation,
 - b. Support independent evaluation of bias, robustness, and safety to ensure that systems are accurately skilled for widespread use,
 - i. Working with accredited labs and cooperating groups to conduct experimentation,
 - ii. It is going to be key to standardize KPIs published quarterly and be actively analyzing and gathering data on false positive rates, appeal outcomes, and redress times,
 - c. Expanding centers that host sandboxes to provide technical support to all states,
 - d. Working along-side public and private companies in enforcing ethical use of Artificial Intelligence;
 - i. Require vendors to perform human rights due diligence aligned with the ethical guidelines for AI usage,

1. Mandating these requirements through AIAs and DPIs for supplied systems;
 - ii. Contractually controlling AI usage through audit rights, software bill of materials, data provenance documentation, security certifications, and any other necessary systems to ensure compliance with basic ethical AI guidelines;
 1. Potentially making use of sanctions and other damages in the case of defiance against ethical codes,
 - e. Specifically ensuring the databases are created on heterogeneous training data, and do not subsequent any discrepancies in representation, especially for marginalised groups;
6. Further requests the launching of a National AI and Digital Rights Literacy and Safety Program coupled with an International AI Literacy Compact to ensure adequate education of the general population, journalists, educators, and officials on AI and specifically spotting AI-enabled manipulation;
 - a. Developing multilingual material AI and rights lessons covering a large variety of information and necessary education
 - i. Implementing material in secondary education, vocational training groups, and adult education in higher education,
 - ii. Focusing on at least privacy and data collection, the function of algorithm systems, consent regarding personal information and property, and recognizing altered-content and disinformation,
 - b. Creating journalist and defender safety tracks that focus on resources and safety information targeted for journalists, human rights defenders, and civil groups,
 - c. Creating community AI clinics and helplines through libraries, schools, and any UN or public infrastructure able to be utilized,
 - d. Providing frontline official training requiring real simulation for training for the police, prosecutors, judges, and public admin on updated rights and respect for the use of AI systems,
 - i. Including specific checklists and information for various protocols and anti-discrimination tactics as well as keeping information up to date and accurate through the UN on the accuracy and potential error-bounds with AI technologies,
 - e. Commit member states to minimum public-education baselines and sharing curriculum modules including peer review and transparency across the UN;

7. Encourages establishment of authenticity infrastructure for synthetic media to protect from AI deception and disinformation while preserving privacy and free expression;
- a. Requiring use of watermarking for any tampered or AI-assisted videos especially for any official governmental statements and communication,
 - i. Creating waterworks that hold strong legal consequences for improper use for MNCs and governmental groups to ensure that videos and information are not AI generated and are from the correct source,
 - b. Launch a public portal for media verification where citizens, journalists, and platforms can check authenticity of official releases and registered public-interest content,
 - c. During election periods, coordinate with electoral authorities and broadcasters allowing for better safety,
 - i. Pre-register official government communications,
 - ii. Provide rapid authenticity checks to newsrooms and run disinformation alerts in collaboration with CSOs,
 - 1. ensure these measures are time-bound and content-neutral;
 - d. Encourage platforms and public agencies to label AI-generated or heavily synthetic content in public-interest contexts with clear, non-sensational badges
 - i. Publish plain-language explanations of terms to ensure accurate and as-intended use in public and online settings.