COMMITTEE: Cyber and AI
QUESTION OF "Setting ethical standards for the use of generative AI technology"
SUBMITTED BY: Germany
CO-SUBMITTED BY: United Kingdom, El Salvador, Japan, USA, France, Hungary, Peru, Senegal, Ukraine, Poland, Italy, Gabon
SIGNATORIES:

The General Assembly,

Recognizing that the rising adoption of generative AI is parallel with issues such as data privacy and protection, distribution of harmful content, and digital security,

Acknowledging that artificial intelligence machines can help advance numerous industries in the future,

Establishing that artificial intelligence developers are well within their rights to create these algorithms independently without any external pressure,

Stressing that trustworthiness is crucial for allowing the diffusion and adoption of AI, therefore allowing digital transformation; and that the public must be well-informed about the beneficial potential of AI,

Affirming that the development of artificial intelligence is a healthy practice and should be continued,

Expressing concern for the damage the misuse of artificial intelligence algorithms can cause as their abilities continue to grow,

Recognizing that the rapid development and adoption of AI calls for set regulations that mitigate the risks derived from this process, promotes a human-centric and trustworthy approach to AI, and fosters research,

1. Establishes an international task force Safe Artificial Intelligence Force (SAIF) as a subset of the UN AI Advisory Board to serve as a platform to discuss, develop, and recommend ethical standards for the use of generative AI technology;
    a. Representatives determined independent of nationality,
        i. Must have at least 10 years experience in cybersecurity, AI development or a related field,
        ii. This includes fostering transparency, conducting regular inspections of accounts, and promoting research and development aimed at eliminating discrimination

       iii. These representatives wll review existing policies to ensure they are adaptable to the landscape of technological advancements,

b. Funding will be provided by the World Bank,

c. SAIF will provide full transparency through the decision making process,
       i. A report shall be filed on the last day of each month outlining the SAIF's decisions,

d. SAIF will support the development of software to tell users whether something is generated by AI or not;
       i. This software will serve as a reliable generative artificial intelligence detection system that will be used worldwide,
       ii. SAIF will promote other research and development opportunities aimed at eliminating discrimination outcomes,

e. Governments may adopt the ethical standards determined by SAIF and implement these policies into national legislation,
       i. Countries **may** request help from SAIF to plan and execute an adoption of SAIF's policies;

2. <u>Introduces</u> the notion of model cards that Artificial Intelligence developers can fill out to inform people of the importance and limitations of their program,

a. These model cards will also provide the context in which each of these algorithms was developed,
       i. Information like the data of development, the intended purpose of the algorithm, any risks the developers notice, and much more can be filled out on these cards,
       ii. These cards also do not need to be fully filled out by developers as they are only aiding in the process while SAIF will be doing the bulk of keeping AI programs that can be used maliciously away from the public,
       iii. Developers may also include new model cards regarding software patches and any new features they are adding to their programs on these model cards,

b. Refusing to submit a model card is within a developer's rights but may warrant,

c. Investigation by the committee depending on the context,

     d.   There can be incentives for developers who choose to submit a model card such as free advertising from the committee which can bolster the revenue the developers are making as it provides them with new consumers,

     e.   SAIF will review the aforementioned model cards;

3. <u>Implores</u> member nations to invest in making AI more accessible by the development of AI in different languages with the intention to create more easily-accessible AI tools for non-native English speaker in numerous languages,

     a.   Some languages these programs could be written in are Italian, Mandarin Chinese, Arabic, French, etc,

     b.   Expanding the development of artificial intelligence worldwide; governments should consider long-term public investments and promote private investment to continue the research and development of trustworthy AI,

     c.   Positioning capable countries to improve their human talent for the research, development, and uses of AI,

          i.   These countries would implement mandatory programming and ethics courses in private and public secondary schools and promote courses in AI for officials of public organizations; digitization and AI jobs;

4. <u>Urges</u> members of government to raise awareness of unethical and illegal problems,

     a.   This is done when considering the use of AI generative art, text or sound by using mediums such as campaigns, advertisements or sponsors,

          i.   These campaigns will target the adolescent and adult demographic,

     b.   If members of government do in fact run advertisements through campaigns the campaign will be funded by high level tech companies,

          i.   SAIF will oversee said campaigns;

5. <u>Stresses</u> the adoption of ethical regulations for sustainable and transparent use of AI with clear definitions of responsibilities and data protection,

     a.   Fostering the development of trustworthy AI by:

          i.   Giving the public a general understanding ofd AI systems,

          ii.   Experimentation and testing of AI systems before becoming public,

        iii.   Making the public aware of their interactions with AI systems,

    b. Preventing biases and discrimination of AI by:

        i.   Ensuring that in the public sector, all cases of use of AI to classify people, data must be kept transparent to the public to guarantee equity,

        ii.   Identifying negative stereotypes and constantly updating and improving software to combat biases using the skills of trained individuals;

6. <u>Encourages</u> reforms in the education sector defined in the digital education strategy on the policy recommendation with the objective to reach teachers and students;

    a. Foster the development of students' general knowledge, abilities, and competencies in AI-related fields in higher education by providing and promoting programs tailored to AI,

    b. Establishing PhD programs in AI and expanding the selection of doctoral courses in AI that are offered,

    c. Providing teacher preparation for supplying educators working in public schools with the expertise needed to assist pupils in utilizing and experimenting with artificial intelligence;

7. <u>Emphasizes</u> the need to increase public knowledge and comprehension of AI in order to set further limiting standards that could lead to the potential increase of jobs,

    a. Aims to improve citizens' digital and AI-related experiences by:

        i.   Educating managers to support the application of AI advancements and broadening their awareness of AI technologies,

        ii.   Educating experts in priority sectors and the public administration,

        iii.   Strengthening the skills of IT and mathematical researchers for both fundamental and applied research,

        iv.   boosting the data engineer competencies' capability growth,

    b. Draws the attention that to create explicit artwork with generative AI, the developers will need a license to do so from the AI company that the creator used.