Guidance on the use of Generative Artificial Intelligence (AI)-based tools in the context of UNEP's seventh edition of the Global Environment Outlook (GEO-7)

Context and background

The United Nations Environment Programme (UNEP) stands at the forefront of global efforts to confront an escalating Triple Planetary Crisis, including climate change, biodiversity loss and pollution. The vast scale and complex nature of the challenges posed by this crisis could benefit from innovative solutions and the application of new digital technologies. In recent years, Artificial Intelligence (AI) has emerged as a transformative technology that can, if used wisely, enhance UNEP's efforts and abilities to help Member States and other stakeholders address these challenges. The UN is monitoring the use of AI and providing recommendations and guidance to be followed. The three most recent and important guidance documents developed by the UN are listed below.

First, the UNESCO Recommendations on the Ethics of Artificial Intelligence were adopted in November 2021. The recommendations provide a global framework for the ethical development and deployment of AI, emphasizing the principles of proportionality and doing no harm, safety and security, fairness and non-discrimination, sustainability, right to privacy and data protection, human oversight and determination, transparency and explainability, responsibility and accountability, awareness and literacy and finally multi-stakeholder and adaptive governance and collaboration. The UNESCO Recommendations on the ethics of AI also include specific provisions addressing the environment and ecosystem impacts of AI.

Second, the UN High-Level Committee on Programmes (HLCP) - Inter-Agency Working Group on Artificial Intelligence issued preliminary operational guidance and a similar series of 10 principles, derived from the UNESCO Recommendations, on the application of AI by the UN System in September 2022. These provide an important set of principles for UNEP's emerging approach.

In addition to these foundational documents, the UN Department of Management Strategy, Policy and Compliance, the UN Department of Field Support and the UN Digital and Technology Network (DTN) have issued internal guidance on the use of generative AI for all UN staff members. The below guidance is consistent with the UN-wide internal guidance issued to staff on 5 July 2023.

Relevance to GEO-7

The resumed fifth session of the UN Environment Assembly (UNEA 5.2) requested UNEP's Executive Director to prepare the seventh edition of the Global Environment Outlook (GEO-7). To ensure the scientific credibility of UNEP's flagship report, the Multidisciplinary Expert Scientific Advisory Group (MESAG) would like to issue guidance to the assessment co-chairs and authors on the use of AI-based Natural Language Processing (NLP) systems and other areas where AI may be used in the GEO process. The increasing use of generative Artificial Intelligence (AI)-based NLP tools, such as OpenAI's generative pre-trained transformer (GPT) model, Reuters-GPT, Climate-GPT or Meta's Llama-2 in scientific and scholarly writing raises important considerations for research ethics and integrity.

While these NLP systems have the potential to enhance scientific article writing and communication, they also come with risks. *The text generated by these tools can mimic human-like writing, but it has become evident that it can contain errors, lack depth, and produce false references or nonsensical connections.* In addition, the use of NLP technology by experts involved in writing 'original works' for UNEP can create **reputational risks for the organization if the narrative and recommendations in a UNEP report are found to have been created by NLP tools rather than the experts themselves**.

This guidance highlights the strengths and weaknesses of NLP systems and provides recommendations for their responsible use to maintain the scientific integrity of the GEO-7 process. **It should be noted that UNEP requires all contributions to GEO-7 to be *original work, with text, research, and graphics solely drafted by the contributing experts.* Therefore, this guidance clarifies how experts can potentially utilize AI tools while ensuring that all contributions to GEO-7 uphold this requirement.**

Generative AI-based Natural Language Processing (NLP)

Natural Language Processing (NLP) enables computers to interact with human language by converting unstructured text into structured text using computational approaches. Previous traditional NLP systems relied on human-written rules, while recent advancements in computational power and machine learning (ML) algorithms, like neural networks, have revolutionized NLP.

Modern AI-based NLP systems employ ML to develop statistical models with billions of parameters. These models are trained on vast databases of text from the internet or other sources, using supervised learning techniques where correct outputs are rewarded by human trainers. Over time, some NLP systems improve their accuracy and learn from the data they process, while others do only one-shot learning in training and then do inference during production. *However, it's important to note that NLP systems do not possess inherent knowledge of the meaning or truth-value of the text. Their purpose is to generate grammatically correct and highly probable text outputs based on the given text inputs.*

Potential Uses of AI-based NLP Systems in Scientific Assessments

AI-based NLP systems are developing rapidly and can potentially interconnect with scientific assessment processes, for example through the following areas:

- Automated Literature Reviews: Conducting comprehensive reviews of scholarly literature on specific topics and time periods. These automated reviews have very high risks of presenting arguments that are not based on established evidence, specifically, references (including sources of literature, data and information) to people, dates, facts, or research in AI-generated content can be entirely fictitious, due to a phenomenon known as “AI hallucination”.
- Language Assistance: Providing suggestions, outside the drafting process for the main scientific assessment documents, for sentence structure, grammar, and vocabulary to improve the clarity and readability of the author-created scientific articles.
- Translation and Multilingual Support: Facilitating communication and understanding across languages by translating articles and reports.
- Question Answering and Information Retrieval: Assisting in finding specific information or answering research-related questions.
- Text Summarization: Analyzing the semantic structure of text to identify key themes and topics.

The above areas are generic, the following sections will discuss strengths and weaknesses of these potential uses followed by specific guidance for the GEO-7 expert community.

Strengths and Weaknesses of using AI-based NLP Systems in Scientific Assessments

Strengths:

- Overcoming Writing Barriers: They can search relevant articles to provide a starting point to overcome writer's block, boost productivity and convert text to academic style. Noting that outputs would be prone to hallucinations and may have unreliable references.
- Complementary Perspective: They can search existing articles for alternative viewpoints, identify missed topics and spark new ideas. The material provided from the search should be verified due to risk of hallucinations and weak references.

- Creative Analogies and Linkages: They can search existing articles to possibly identify lateral, non-obvious, connections between concepts, though a reality check is necessary.
- Manuscript Refinement: They can suggest improvements to titles, abstracts, conclusions, and overall structure, which should be counterchecked for accuracy in how the main text has been summarized.
- Enhanced Literature Search: They can search for relevant references that may be missed in conventional searches; however, the results of the search will need to be reviewed for quality and relevance of references as well as hallucinations.
- Writing Structure Guidance: They can search relevant articles to simplify complex topics for better understanding and organization of concepts and narratives. The results will be prone to hallucinations with a risk of unreliable references.
- Inclusivity for Non-Native English Speakers: They can provide fairly accurate translations of articles and craft summaries of these, thereby supporting non-native English speakers in their writing process. Translations should be counterchecked for quality control.
- Comprehensive Coverage: They can search relevant articles to encourage consideration of overlooked aspects, with risks of hallucinations and unreliable references.
- Knowledge Expansion: They can structure information in a simplified way, allowing better understanding of concepts, but will lack creative insights and may misconstrue the meaning of the original text.

Weaknesses:

- Errors and Inaccuracies: NLP-generated text can contain errors, inaccuracies, and superficial content. For example: hallucinations, inaccurate or irrelevant references, etc.
- Lack of Critical Thinking: NLP tools do not possess critical thinking abilities and may make nonsensical or false connections.
- Contextual Understanding: They will not grasp the meaning and context of complex scientific concepts, leading to some potentially problematic results (i.e. hallucinations).
- Limited Creativity: NLP-generated text may lack originality, creative insights and innovative ideas.
- Lack of Domain Expertise: NLP models have limitations in specialized domains and will not capture the nuances of specific scientific fields, including domain-specific language, expressions, acronyms or similar.
- Ethical Concerns: Unethical use of NLP tools can lead to the creation of fake or low-quality scientific content.
- Reliance on Existing Data with Risks of Bias: NLP models heavily rely on their training data, which comes with issues such as limited time series, violation of intellectual property rights, hallucinations and risks of bias especially in domains that rely heavily on judgement and have a wide range of opinions.
- Data Extraction and Analysis: NLP systems cannot extract and analyse data without being combined with other types of models, which can generate inaccurate results which renders the outputs unusable in best practice scientific assessments.

Guidelines for using AI-based NLP Systems in GEO-7 Assessments

When using NLP systems in scientific research and manuscript writing, the following considerations should be taken into account:

- Verification by Domain Experts: Literature searches and analysis generated by NLP systems should be thoroughly checked by domain experts to ensure accuracy, relevance, absence of bias, and logical reasoning.

- Author Responsibility: Authors are ultimately responsible for producing all text contained in the final manuscript (e.g., GEO-7 report) and should be held accountable for any inaccuracies, fallacies, or problems that may arise from the use of NLP tools.
- Research and Analysis: Authors should transparently disclose their use of NLP systems and clearly indicate which research, analysis or data were obtained through the use of NLP tools, ensuring readers have a complete understanding of the supporting analysis in the text produced by the author/expert.
- Data Integrity: Researchers should refrain from using NLP systems to fabricate empirical data or falsify existing data, as it violates various codes of ethics and undermines the integrity of research supporting the analysis conducted by the author.
- Impact on Content: No direct use of NLP-generated text should be included in any manuscript or report produced for UNEP. Any influence of NLP assistance on the text produced by the author for the publication should be disclosed to maintain transparency and prevent potential questions of scientific integrity or legitimacy related to the publication.
- Any use of Generative AI/NLP systems in the GEO-7 assessment is subject to prior approval by UNEP and public disclosure in each publication.

Adhering to these guidelines will contribute to safeguarding the scientific credibility of the GEO-7 assessment and avoid any ethical violations from other UN guidance. New technologies and tools, under the current rate of expansion and development, do present potential opportunities and risks that GEO-7, and the scientific community as a whole, should continue to monitor and document.

Additional considerations: video and audio avatars

Various AI tools can be used to create video and audio avatars of individuals that are realistic enough so as to be indistinguishable from the person or persons in question. Through the use and recording of virtual meetings during the GEO-7 process, large numbers of video and audio likenesses of the GEO-7 participants are stored on UNEP and Microsoft servers. These likenesses could be retrieved through a hack of these systems and AI tools could be used in malicious ways to impersonate some of the experts involved in the process. UNEP has a responsibility to put in place measures to minimize the likelihood of this malicious use happening.

During GEO-7 processes, the video and audio recording of participants should be used as infrequently as possible. The recording of transcripts from calls should be the main way of ensuring decisions and discussions are recorded, for the benefit of other participants who were unable to attend the call and for use by UNEP to accurately document the proceedings from these virtual or hybrid meetings.

10 August 2023

Annex 1: UN Inter-Agency Working Group on Artificial Intelligence 10 principles:

1. Do no harm: AI systems should not be used in ways that cause or exacerbate harm, whether individual or collective, and including harm to social, cultural, economic, natural, and political environments. All stages of an AI system lifecycle should operate in accordance with the purposes, principles and commitments of the Charter of the United Nations. All stages of an AI system lifecycle should be designed, developed, deployed and operated in ways that respect, protect and promote human rights and fundamental freedoms. The intended and unintended impact of AI systems, at any stage in their lifecycle, should be monitored in order to avoid causing or contributing to harm, including violations of human rights and fundamental freedoms.

2. Defined purpose, necessity and proportionality: The use of AI systems, including the specific AI method(s) employed, should be justified, appropriate in the context and not exceed what is necessary and proportionate to achieve legitimate aims that are in accordance with each United Nations system organization's mandates and their respective governing instruments, rules, regulations and procedures.

3. Safety and security: Safety and security risks should be identified, addressed and mitigated throughout the AI system lifecycle to prevent where possible, and/or limit, any potential or actual harm to humans, the environment and ecosystems. Safe and secure AI systems should be enabled through robust frameworks.

4. Fairness and nondiscrimination: United Nations system organizations should aim to promote fairness to ensure the equal and just distribution of the benefits, risks and costs, and to prevent bias, discrimination and stigmatization of any kind, in compliance with international law. AI systems should not lead to individuals being deceived or unjustifiably impaired in their human rights and fundamental freedoms.

5. Sustainability: Any use of AI should aim to promote environmental, economic and social sustainability. To this end, the human, social, cultural, political, economic and environmental impacts of AI technologies should continuously be assessed and appropriate mitigation and/or prevention measures should be taken to address adverse impacts, including on future generations. These may include, for example, the development or enhancement of a) sustainable, privacy-protected data access frameworks, b) appropriate safeguards against function creep, and c) fair and inclusive training, validation, and maintenance of AI models utilizing quality data.

6. Right to privacy, data protection and data governance: Privacy of individuals and their rights as data subjects must be respected, protected and promoted throughout the lifecycle of AI systems. When considering the use of AI systems, adequate data protection frameworks and data governance mechanisms should be established or enhanced in line with the United Nations Personal Data Protection and Privacy Principles also to ensure the integrity of the data used.

7. Human autonomy and oversight: United Nations system organizations should ensure that AI systems do not overrule freedom and autonomy of human beings and should guarantee human oversight. All stages of the AI system lifecycle should follow and incorporate humancentric design practices and leave meaningful opportunity for human decision-making. Human oversight must ensure human capability to oversee the overall activity of the AI system and the ability to decide when and how to use the system in any particular situation, including whether to use an AI system and the ability to override a decision made by a system. As a rule, life and death decisions or other decisions affecting fundamental human rights of individuals must not be ceded to AI systems, as these decisions require human intervention.

8. Transparency and explainability: United Nations system organizations should ensure transparency and explainability of AI systems that they use at all stages of their lifecycle and of decision-making processes involving AI systems. Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. Individuals should be meaningfully informed when a decision which may or will impact their rights, fundamental freedoms, entitlements, services or benefits, is informed by or made based on AI algorithms and have access to the reasons for a decision and the logic involved. The information and reasons for a decision should be presented in a manner that is understandable to them.

9. Responsibility and accountability: United Nations system organizations should have in place appropriate oversight, impact assessment, audit and due diligence mechanisms, including whistle-blowers' protection, to ensure accountability for the impacts of the use of AI systems throughout their lifecycle. Appropriate governance structures should be established or enhanced which attribute the ethical and legal responsibility and accountability for AIbased decisions to humans or legal entities, at any stage of the AI system's lifecycle. Harms caused by and/or through AI systems should be investigated and appropriate action taken in response. Accountability mechanisms should be communicated broadly throughout the United Nations system in order to build shared knowledge resources and capacities.

10. Inclusion and participation: When designing, deploying and using AI systems, United Nations system organizations should take an inclusive, interdisciplinary and participatory approach, which promotes gender equality. They should conduct meaningful consultations with all relevant stakeholders and affected communities, in the process of defining the purpose of the AI system, identifying the underlying assumptions, determining the benefits, risks, harms and adverse impacts, and adopting prevention and mitigation measures.