



# Integrating artificial intelligence with expert knowledge in global environmental assessments: opportunities, challenges and the way ahead

Veruska Muccione<sup>1,2</sup> · Saeid Ashraf Vaghefi<sup>1,3,4</sup> · Julia Bingler<sup>5,6</sup> · Simon K. Allen<sup>1,7</sup> · Mathias Kraus<sup>8</sup> · Glen Gostlow<sup>3</sup> · Tobias Wekhof<sup>3,9</sup> · Chiara Colesanti-Senni<sup>3</sup> · Dominik Stambach<sup>10</sup> · Jingwei Ni<sup>3,9</sup> · Tobias Schimanski<sup>3</sup> · Tingyu Yu<sup>3</sup> · Qian Wang<sup>3</sup> · Christian Huggel<sup>1</sup> · Juerg Luterbacher<sup>4</sup> · Robbert Biesbroek<sup>11</sup> · Markus Leippold<sup>3,12</sup>

Received: 19 February 2024 / Accepted: 21 July 2024  
© The Author(s) 2024

## Abstract

With new cycles of global environmental assessments (GEAs) recently starting, including GEO-7 and IPCC AR7, there is increasing need for artificial intelligence (AI) to support in synthesising the rapidly growing body of evidence for authors and users of these assessments. In this article, we explore recent advances in AI and connect them to the different stages of GEAs showing how some processes can be automatised and streamlined. The meticulous and labour-intensive nature of GEAs serves as both a valuable strength and a challenge to staying pertinent and current in today's era of urgency and the pursuit of the latest knowledge. Utilising AI tools for reviewing and synthesizing scientific literature holds the evident promise of substantially lessening the workload for experts and expediting the assessment process. This, in turn, could lead to more frequent report releases and a smoother integration of the latest scientific advancements into actionable measures. However, successful outcomes can only be achieved if domain experts co-develop and oversee the deployment of such tools together with AI researchers. Otherwise, these tools run the risk of producing inaccurate, incomplete, or misleading information with significant consequences. We demonstrate this through a few examples that compare recently deployed large language models (LLMs) based tools in their performance in capturing nuanced concepts in the context of the reports of the Intergovernmental Panel on Climate Change (IPCC). We recommend establishing ethical committees and organising dedicated expert meetings to develop best practice guidelines, ensuring responsible and transparent integration of AI into GEAs.

**Keywords** Global environmental assessments · Climate change · Artificial intelligence · Large language models

## Introduction

Global environmental assessments (GEAs) such as those developed by the Intergovernmental Panel on Climate Change (IPCC), the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), Global Environmental Outlook (GEO), the Global Sustainable Development Report, and UNEP Emission Gap and Adaptation Gap reports are crucial to provide information and foster a better understanding of the causes and consequences of natural and human activities on the environment and to develop action-oriented and effective solutions which promote planetary health and wellbeing (Castree et al. 2020). In the past, these

assessments have informed key policy decisions and helped to identify gaps in current environmental governance frameworks (Jabbour and Flachsland 2017). The IPCC for example has a long history of informing national and international policy and negotiations, for instance through the United Nations Framework Convention on Climate Change (UNFCCC). The panel's recommendations have underpinned the development of national and international climate policies, including the Paris Agreement (Ourbak and Tubiana 2017). More recently, the reports of the 6th IPCC Assessment cycle (thereafter IPCC AR6) have played a crucial role in raising public awareness about the urgent need to address climate change (IPCC 2023) and informing the Global Stocktake on achieving mitigation, adaptation and means of implementation targets. A few concerns though have been raised for GEAs (Vardy et al. 2017; Castree et al. 2020), not least the exponentially growing body of evidence from the peer-reviewed and grey literature

---

Communicated by Wolfgang Cramer

Extended author information available on the last page of the article

(Stocker and Plattner 2014; Berrang-Ford et al. 2021a; Palutikof et al. 2023). A bibliometric analysis of the Web of Science has shown that the number of unique articles on climate change has grown from almost 40,000 between 2010 and 2015 to more than 57,000 between 2016 and 2021. During this time, there has also been a clear shift from the physical science of climate change to the topics of impacts, mitigation and adaptation (Khojasteh et al. 2024). Consequently, the number of references that have been used in each IPCC assessment cycle has also rapidly increased. For example, in the 5th assessment cycle, the IPCC authors assessed approximately 30,000 publications,<sup>1</sup> whereas in the 6th assessment more than 66,000 publications have been assessed.<sup>2</sup> Not only the number of references has increased, but over the years, the scope of GEAs has become more diverse. Today's environmental assessments draw on a much wider range of fields, including natural and social sciences, economics, anthropology, psychology, engineering and the humanities (Callaghan et al. 2020). Similarly, the readership of these reports has increased, from mostly negotiators and national governments in early assessments, to a range of (non)-governmental organisations, civil society, and businesses across the globe.

Given the growth in environmental literature and the expanding scope and readership of global GEAs, traditional methods of synthesizing work are increasingly challenging.

## Artificial intelligence assisted synthesis work

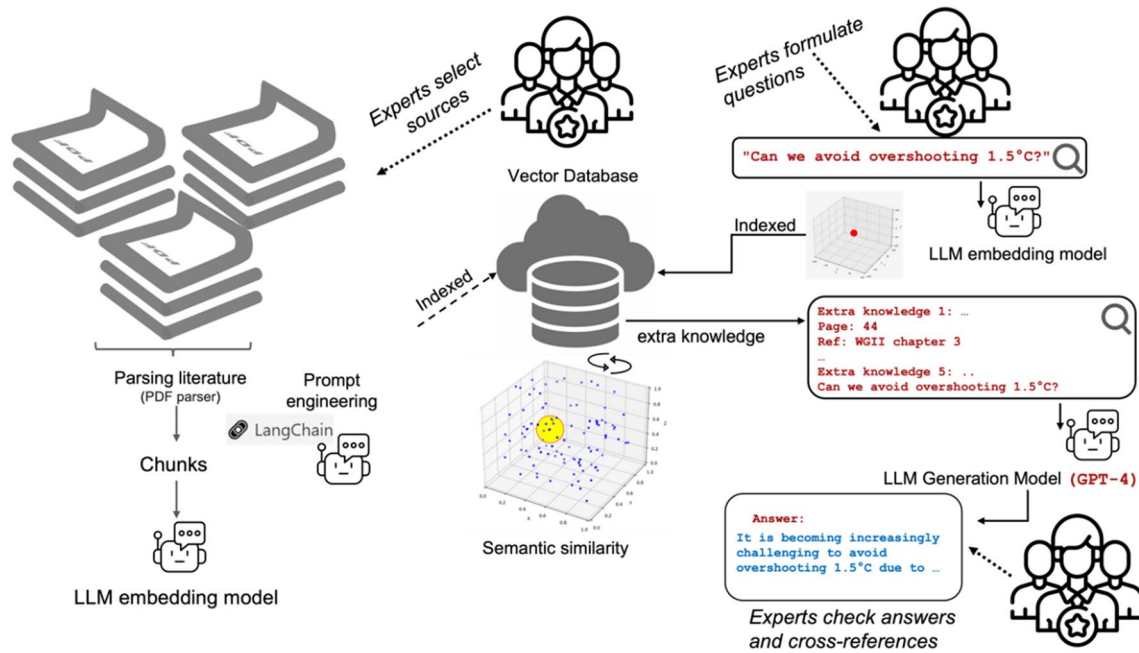
Assessing the evidence from many diverse sources and disciplines is an incredibly demanding task. The authors need to search, collect and systematically assess the literature to synthesise evidence and extract coherent narratives that can be traced back to their original sources (Callaghan et al. 2020; Berrang-Ford et al. 2021a). At the same time, machine learning (ML)-assisted research synthesis has grown in popularity amongst the environmental research community (Callaghan et al. 2020; Berrang-Ford et al. 2021a; Lydiri et al. 2022; Sietsma et al. 2024a). Advances in natural language processing (NLP) have led to pre-trained large language models (LLMs), which are currently standard in the field.

(Brown et al. 2020). In the climate domain, studies have used transformers based LLMs from the BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) family for analysing corporate climate disclosure (Bingler et al. 2022) and for climate impact attribution (Callaghan

et al. 2021). Some machine learning methods have also helped in the IPCC AR6 cycle. For example, supervised and unsupervised learning techniques have been used to assess the evidence of human adaptation to climate change (Berrang-Ford et al. 2021a), the extent of climate change health related topics (Berrang-Ford et al. 2021b) and the existence of adaptation limits (Thomas et al. 2021). Similarly, these methods allow for multilingual assessments and automated translation of texts, which addresses a frequently heard critique in global assessments namely the reporting bias towards English language evidence (Sietsma et al. 2023). For example, machine translation (MT) can be integrated at the system level to facilitate multi-language support. In this process, non-English queries are initially translated into English to allow the LLM engine to produce an English response. Subsequently, this response can be translated back into the original language (Thulke et al. 2024). Artificial intelligence (AI) supported screening of the rapidly increasing body of literature can save significant time and scarce resources (van de Schoot et al. 2021) as well as creating “living evidence synthesis platforms” that allow for continuously updating the expanding pool of scientific literature to be included in the assessment (Elliott et al. 2021). One way to do this is by using AI Research Assistant tools to survey the literature prior to evidence synthesis which allows to find papers, extract data and arrange them in concepts (De-Gol et al. 2023). In a next stage, AI tools can assist authors in determining the level of confidence in their statements by more systematically and transparently characterising the level of agreement and the level of evidence (Mastrandrea et al. 2011). To pursue this, authors can adopt a fact-checking approach (Leippold et al. 2024). This architecture enables LLMs to proficiently integrate diverse scientific evidence and iteratively use a mediator-advocate framework to converge towards a final assessment based on multiple lines of evidence. This somehow mirrors a human expert judgement elicitation, which is often at the core of such assessments (Majszak and Jebeile 2023). Fact checking can also go a step further and be used to check for consistency across different elements of a report, as for example amongst the chapters and to illuminate about potential knowledge gaps. These advances in AI have led to vastly increased usage of text-as-data methods in the context of environmental and climate change and beyond, which so far has only received limited attention (Stede and Patz 2021). In the next sections, we critically discuss the capabilities, opportunities and usefulness of question answering tasks (QA) following the emergences of tools deployed in 2023, primarily ChatClimate (Vaghefi et al. 2023) which was developed by a group of authors of this paper, but for completeness also ClimateQA (Lelong et al. 2023) and ClimateGPT (Thulke et al. 2024).

<sup>1</sup> [https://archive.ipcc.ch/news\\_and\\_events/docs/ar5/150318\\_SYR\\_final\\_publication\\_pr.pdf](https://archive.ipcc.ch/news_and_events/docs/ar5/150318_SYR_final_publication_pr.pdf)

<sup>2</sup> [https://www.ipcc.ch/site/assets/uploads/2022/02/AR6\\_Factsheet\\_2022\\_February.pdf](https://www.ipcc.ch/site/assets/uploads/2022/02/AR6_Factsheet_2022_February.pdf)



**Fig. 1** QA-Experts workflow. The QA-Experts workflow starts with experts adding extra (high quality and trusted) knowledge, such as IPCC and/or other GEA reports and peer-reviewed articles to the database (left-hand side of the panel). Then, experts formulate the appropriate questions from the body of literature and feed questions

to the chatbots (left-hand side of the panel). Based on a semantic search, related texts to each question are retrieved through a contextual compression system (middle part and right middle). The experts finally check the answers and check the references within (bottom right) (adapted from Vaghefi et al. 2023)

### Leveraging the benefits

Question Answering (QA) involves using natural language processing (NLP) techniques to develop models that can understand and respond to questions posed in natural language (Van Dis et al. 2023). These models are trained on large datasets and can identify and extract relevant information to provide an answer to a user’s question. Powerful LLMs trained for QA can comprehend complex questions, interpret their underlying meaning and context, and apply the knowledge they have acquired during training to generate accurate and informative answers to a query. LLMs can be further extended to chatbots, offering users an interactive and intelligent dialogue in a contextually relevant manner (Stokel-Walker and Van Noorden 2023). They are estimated to have considerable economic, social and policy implications (Eloundou et al. 2023). However, they are subject to two major challenges: hallucination (i.e. generation of text that is not grounded in factual information) and outdated information that only dates to the time of training (Leeming 2023). In domains like global change, getting accurate, up-to-date information from trustworthy sources is essential. A possible solution has been to (i) give LLMs access to additional, scientifically acknowledged resources to keep their knowledge up to date and avoid spreading disproven, outdated or misleading information, and (ii) ask

a wide range of experts to check the reliability of answers (Gao et al. 2023).

Following this rationale, domain-specific chatbots like ChatClimate, ClimateQA and ClimateGPT have been deployed to enhance LLMs, through the integration of information from databases of relevant documents, namely the IPCC Assessment Reports, the IPBES and documents of well-established organisations such as the World Meteorological Organisation (WMO) (Vaghefi et al. 2023). These chatbots can be utilised in a workflow jointly or in tandem designed with climate researchers, streamlining the process of adding literature to databases and accelerating literature search, collection and cross-referencing efforts. This is illustrated in Fig. 1 for the specific case of ChatClimate. A detailed technical description of the chatbot is beyond the scope and the reader is referred to the overview by Vaghefi et al. (2023).

To ensure that the scientific community remains effective in synthesizing and timely conveying knowledge, leveraging this QA capabilities of chatbot built upon LLMs in support of authors and report beneficiaries can be particularly useful. For authors, this is a help with respect to reports relying on very diverse literature, stemming from various disciplines. For users of the reports, these systems can play a role in extracting and processing information tailored to their questions. Chatbot tools based on curated corpora can be both

innovative and powerful with respect to the types of questions end-users of reports want answers to, allowing reports to be scoped in a way that not only better addresses the needs of decision-makers, but also that of other end-users. As illustrative example, we perform a semantic analysis of the three assessment reports of the IPCC AR6 (ESM Fig. 1, left panel) and of the questions asked to ChatClimate between April and November 2024 (ESM Fig. 1, right panel). By means of world clouds, we can see that there is reasonable agreement in terms of the most common words in the reports and the questions asked to that report. For example, in both cases, the word “risk”, “emission” or “warming” are present with similar frequency. Notwithstanding, we notice how the words “impacts” or “adaptation” are hardly visible in the questions world cloud despite being very prevalent in the IPCC reports. A fairly simple analysis of the questions across all tools can help steering, for example outreach materials after publication, targeting the broader community and customizing content for different user categories. A community of users can be created which interact with authors through specific chatbots to elicit aspects of the report in full spirit of knowledge co-creation (Mauser et al. 2013). Additionally, LLMs can be used for factchecking to counteract the spreading of misinformation, i.e. to check the validity of claims made through various media and social media platforms (Chavalarias 2022; Leippold et al. 2024; Schimanski et al. 2024).

## Testing response accuracy

In this section, we demonstrate the performance and limitations of using chatbots through practical experiments, utilising the IPCC reports as a testbed for these technologies. These experiments can of course also be adapted accordingly to other GEAs.

In a first experiment, three different instructions are provided to ChatClimate on how to answer a query. As outlined in the previous section, a crucial step is the integration of domain-specific material on top of the general and large quantity of text data that it is provided to LLMs during training (Fig. 1). As it can be seen in Table 1, the same question is asked to the underlying LLM (GPT-4), but only in the *hybrid* and *standalone* cases the model is fed with additional information from the IPCC reports. Furthermore, prompt engineering is used in the *standalone* case to force the model to only base its answers on the IPCC reports. It can be noticed that the answer returned using only GPT-4 is seemingly correct, but not necessarily focused and lacks the nuances and details that are more discernible in the other two cases. Furthermore, the use of in text citations with possibility to retrieve the page number from the *hybrid* and

*standalone* models is certainly an advantage as it allows to trickle back statements to the source of information.

The second experiment focuses on the ability of LLMs to capture nuanced concepts such as *accelerating climate change* and *tipping points*. In this experiment, ChatClimate was first given only the AR6 Synthesis Report (IPCC AR6 SYR) to formulate its answer, and then, it was given all underlying IPCC Sixth Assessment Reports but excluding the IPCC AR6 SYR, to answer the same question. In the supplementary material, ESM Table 1 shows the answers of ChatClimate when using only the IPCC AR6 SYR versus the case where all reports except the IPCC AR6 SYR are included. The answers to the question of acceleration are evaluated by the authors as being inaccurate in both cases. ChatClimate seems to confuse the concept of acceleration with “rate of change”. An acceleration is mathematically distinct from a rate of change and the only conclusion in the synthesis report on acceleration is accelerating sea level rise. Hence, a limited mathematical characterisation of “acceleration” in the report (except for evidence on accelerating sea-level rise) might have given rise to the confusion. The answer to the second question on tipping points shows that using only the IPCC AR6 SYR leads to a more accurate response than the case where it is excluded. This experiment thus highlights the value of iterative learning when preparing the data source and the importance to involve the authors who have actually produced the data source, particularly when it comes to concepts that have multiple meanings.

In the third and last experiment, the answers provided by the three different tools ChatClimate, ClimateGPT and ClimateQA were compared. Again, we asked a question which requires accuracy. The chatbots are asked *how much of climate change is due to fossil fuels*. The results are summarised in ESM Table 2. Although different responses styles and level of details are given, in this case only ClimateQ&A reflects the full magnitude of CO<sub>2</sub> emissions from fossil fuels. ChatClimate only focuses on the magnitude of the past decade.

These three short experiments call for a in depth reflection, evaluation as well as comparison of tools from the side of domain experts such as climate scientists. A compilation of sensitive and nuanced topics should become a standard procedure in the age of AI as well as a comparison on how such topics are treated across different elements of the reports. Hence, it is essential that experts who are familiar with the breadth and scope of each specific topic test these tools regularly. In absence of regular testing and evaluation, potentially well-crafted and convincing answers might hide incorrect or outdated statements (Bender et al. 2021). The process could be improved by performing frequent rounds of response ranking that engages a large pool of researchers, covering the full range of scientific views as well as LLMs serving as advocates (Leippold et al. 2024).

**Table 1** The table compares the three different model setups, namely GPT-4 is only using knowledge from the training material (GPT-4), GPT-4 is instructed to use IPCC reports on top of the in-house knowl- edge (GPT-4 + ChatClimate) and finally in the standalone set-up GPT-4 is instructed to only use the IPCC reports to answer the query (Standalone ChatClimate)

Query: “Why there is a need to phase out fossil fuels to keep global temperature increase below 1.5 °C?”			
	GPT-4	Hybrid (GPT4 + ChatClimate)	Standalone ChatClimate
Experiment details	No extra information or instruction is given to the model on how to answer the questions	The model is instructed to provide answers based on the IPCC reports (in this example the IPCC Synthesis report only) and the general in-house knowledge of the Open AI model GPT-4. The model is also instructed to give priority to the IPCC report	The model is instructed to provide answers based exclusively on the IPCC reports (in this example the IPCC Synthesis report only)
Response to the query	Fossil fuels such as coal, oil and natural gas are the main sources of human-caused greenhouse gas emissions, contributing to about 75% of the world’s total emissions. When burned, these fuels release carbon dioxide and other greenhouse gases into the atmosphere, which trap heat from the sun and cause the planet to warm—a phenomenon known as global warming. To keep the global temperature, increase below 1.5 °C—a target set by the Paris Agreement to prevent the worst impacts of climate change—it’s necessary to drastically reduce these emissions. This means phasing out the use of fossil fuels and transitioning to cleaner, renewable sources of energy like wind, solar and hydro power. If we continue to burn fossil fuels at our current rate, we’re likely to exceed this 1.5 °C target, leading to more severe and frequent heatwaves, storms, flooding and other extreme weather events. It could also cause sea levels to rise, threatening coastal communities, and lead to a loss of biodiversity as habitats change or disappear. Therefore, phasing out fossil fuels is crucial for mitigating these potential impacts of climate change	Phasing out fossil fuels is necessary to keep the global temperature increase below 1.5 °C because the cumulative CO2 emissions from burning fossil fuels contribute significantly to global warming. According to the IPCC reports, limiting human-caused global warming to a specific level requires limiting cumulative CO2 emissions, reaching net zero CO2 emissions, and making strong reductions in other greenhouse gas emissions (Page 50, IPCC-SYR). The projected CO2 emissions from existing fossil fuel infrastructure without additional abatement would exceed the remaining carbon budget for 1.5 °C (Page 25, IPCC-SYR). This means that if we continue to burn fossil fuels at the current rate, we will emit more CO2 than the amount that can be absorbed by natural processes, leading to an increase in global temperatures beyond 1.5 °C Furthermore, about 80% of coal, 50% of gas and 30% of oil reserves cannot be burned and emitted if warming is to be limited to 2 °C. Significantly more reserves are expected to remain unburned if warming is limited to 1.5 °C (Page 24, IPCC-SYR). This indicates that a significant portion of the world’s fossil fuel reserves must remain unexploited to prevent excessive global warming Finally, achieving and sustaining net negative global CO2 emissions, which would gradually reduce the warming level, requires the deployment of Carbon Dioxide Removal (CDR) technologies. These technologies are necessary to counterbalance hard-to-abate residual emissions from sectors like agriculture, aviation, shipping and industrial processes (Page 50, IPCC-SYR). However, the most effective way to reduce CO2 emissions is to phase out the use of fossil fuels This information is based on the IPCC AR6 reports	Phasing out fossil fuels is necessary to keep the global temperature increase below 1.5 °C because the cumulative carbon emissions from existing fossil fuel infrastructure, without additional abatement, would exceed the remaining carbon budget for 1.5 °C (Page 25, IPCC-SYR). This means that the CO2 emissions from burning fossil fuels would push global warming beyond the 1.5 °C target Moreover, limiting global warming to a specific level requires limiting cumulative CO2 emissions and reaching net zero CO2 emissions, along with strong reductions in other greenhouse gases (Page 46, IPCC-SYR). This implies that we need to stop adding more CO2 to the atmosphere, which is primarily released by burning fossil fuels Furthermore, about 80% of coal, 50% of gas and 30% of oil reserves cannot be burned and emitted if warming is to be limited to 2 °C, and significantly more reserves are expected to remain unburned if warming is limited to 1.5 °C (Page 24, IPCC-SYR). This indicates that a large portion of the existing fossil fuel reserves must remain unexploited to avoid exceeding the 1.5 °C target Finally, achieving and sustaining net negative global CO2 emissions, which would gradually reduce the warming level, requires the deployment of Carbon Dioxide Removal (CDR) to counterbalance hard-to-abate residual emissions, such as those from fossil fuels (Page 50, IPCC-SYR). This further emphasises the need to phase out fossil fuels and transition to cleaner energy sources

Convergence in response rankings could be elicited following a Delphi style procedure such as the one performed in IPCC AR6 (Zommers et al. 2020). Moreover, the performance of the technology can be improved by using sensibly curated corpora. These corpora can in principle be tailored to specific topics as, for example adaptation (Sietsma et al. 2024a, b), where using AI allows for small teams of authors (e.g. authors of IPCC special reports or chapter authors) to develop queries tailored to the information they are interested in (De-Gol et al. 2023). In the context of the IPCC, the Technical Support Unit or another initiative (e.g. as was done with the Global Adaptation Mapping Initiative, see Berrang-Ford et al. 2021a) may play a role in this by providing authors with the necessary skillset and resources needed to take full advantage of the AI capability.

### The way forward in the age of AI

Deploying domain specific LLMs provides researchers and users of scientific knowledge targeted access to specific information and answers. Using LLMs and specialised chatbots could relieve authors from the lengthy searching and meticulous procedures of reading through thousands of papers and leave more time for synthesis work which include compiling and evaluating the evidence (Mach et al. 2017). Using AI tools in the literature selection and review process also minimises potential subjectivity in the assessment, thereby ensuring these global assessments can continue to assess the full and diverse range of scientific perspectives. It is suggested that the scientific community can greatly benefit from the use of AI for better scientific communication, knowledge accessibility and

synthesis. In line with open science and open research data principles, researchers from around the world can access and contribute to shared knowledge, promoting a more inclusive and globally connected community. AI has clear potential to improve the efficiency of the process and ultimately enhance the comprehensiveness and usability of the reports.

The careful and labour-intensive process of Global Environmental Assessments has proven to be both its greatest asset, but also a barrier to remaining relevant and up to date in the modern age of urgency and thrust for latest knowledge. The use of AI tools to assist in the review and synthesis of the scientific literature can reduce the burden on experts and expedite the assessment process, thereby paving the way towards more regular release of reports and streamlining the transfer of latest science into action. So far, we have not addressed the capability of large language models (LLMs) to interpret figures and tables, but this is becoming increasingly feasible. While it remains a challenge to fully interpret the complex and rich figures found in GEA reports, the rapid pace of advancements suggests that this capability may become standard practice sooner rather than later (OpenAI 2024).

Clearly, AI tools are not a substitute for the rigour that only experts can provide. As it is demonstrated in the previous section, there are nuanced concepts which might be tangentially covered in GEAs or for which there is limited evidence and scientific consensus. A periodic evaluation of the tools with a large pool of experts must be encouraged to secure that answers are anchored in latest science and to avoid misleading information. Given the rapidity with which AI technologies are developing, it is essential to establish ethical procedures for their use. This involves addressing ethical considerations such as reproducibility and bias mitigation. We recommend the creation of ethical committees, and the organisation of dedicated expert meetings. Specifically, we call for the development of guidelines for best practices in integrating AI into global environmental assessments, ensuring that the adoption of these technologies is both responsible and transparent. Additionally, training LLM models on huge amounts of data has a potentially very high carbon footprint and we have little knowledge about the carbon footprint embedded in LLMs such as GPT-4 (Bender et al. 2021). Inference, and the use of already trained LLM models, is increasingly energy intensive, the larger the models are. Therefore, the LLMs community itself needs to implement environmentally aware workflows to avoid contributing to the challenges they claim to tackle (Hershcovich et al. 2022).

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10113-024-02283-8>.

**Acknowledgements** V.M acknowledges support of the Digitalization Initiative of the Zurich Higher Education Institutions (DIZH). V.M would like to thank G.K.P for his support during the preparation of this manuscript. The authors would like to thank the reviewers for their time and invaluable recommendations.

**Funding** Open access funding provided by University of Zurich.

## Declarations

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.


## References

- Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: Can language models be too big? In Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10. Canada. Association for Computing Machinery Inc, Virtual Event, pp 610–623. <https://doi.org/10.1145/3442188.3445922>
- Berrang-Ford L, Siders AR, Lesnikowski A, Fischer AP, Callaghan MW, et al. (2021a) A systematic global stocktake of evidence on human adaptation to climate change. *Nat Clim Change* 11:989–1000. <https://doi.org/10.1038/s41558-021-01170-y>
- Berrang-Ford L, Sietsma AJ, Callaghan M, Minx JC, Scheelbeek PFD et al (2021b) Systematic mapping of global research on climate and health: a machine learning review. *Lancet Planet Health* 5:e514–e525. [https://doi.org/10.1016/S2542-5196\(21\)00179-0](https://doi.org/10.1016/S2542-5196(21)00179-0)
- Bingler JA, Kraus M, Leippold M, Webersinke N (2022) Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures. *Finance Res Lett* 47:102776–102776. <https://doi.org/10.1016/j.frl.2022.102776>
- Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, et al (2020) Language Models are Few-Shot Learners. *arXiv:2005.14165*. <https://doi.org/10.48550/arXiv.2005.14165>
- Callaghan MW, Minx JC, Forster PM (2020) A topography of climate change research. *Nat Clim Change* 10:118–123. <https://doi.org/10.1038/s41558-019-0684-5>
- Callaghan M, Schleussner CF, Nath S, Lejeune Q, Knutson TR et al (2021) Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nat Clim Change* 11:966–972. <https://doi.org/10.1038/s41558-021-01168-6>
- Castree N, Bellamy R, Osaka S (2020) The future of global environmental assessments: Making a case for fundamental change. *Anthr Rev* 8:56–82. <https://doi.org/10.1177/2053019620971664>
- Chavalarias D (2022) TOXIC DATA - Comment les réseaux manipulateurs nos opinions. Flammarion
- De-Gol AJ, Le Quéré C, Smith AJP, Le Quéré MA (2023) Broadening scientific engagement and inclusivity in IPCC reports through collaborative technology platforms. *Npj Clim Action* 2:49–49. <https://doi.org/10.1038/s44168-023-00072-3>

- Elliott J, Lawrence R, Minx JC, Oladapo OT, Ravaud P et al (2021) Decision makers need 'living' evidence synthesis. *Nature* 600:383–385
- Eloundou T, Manning S, Mishkin P, Rock D (2023) GPTs are GPTs: an early look at the labor market impact potential of large language models. *arXiv:2303.10130*. <https://doi.org/10.48550/arXiv.2303.10130>
- Gao Y, Xiong Y, Gao X, Jia K, Pan J, et al (2023) Retrieval-augmented generation for large language models: a survey. *arXiv:2312.10997*. <https://doi.org/10.48550/arXiv.2312.10997>
- Hershcovich D, Webersinke N, Kraus M, Bingler JA, Leippold M, et al (2022) Towards climate awareness in NLP research. *arXiv:2205.05071*. <https://doi.org/10.48550/arXiv.2205.05071>
- IPCC (2023) Summary for policymakers. In: *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 1–34. <https://doi.org/10.59327/IPCC/AR6-9789291691647.001>
- Jabour J, Flachslund C (2017) 40 years of global environmental assessments: a retrospective analysis. *Environ Sci Policy* 77:193–202. <https://doi.org/10.1016/j.envsci.2017.05.001>
- Khojasteh D, Haghani M, Shamsipour A, Zwack CC, Glamore W et al (2024) Climate change science is evolving toward adaptation and mitigation solutions. *Wires Clim Change* 15:e884–e884. <https://doi.org/10.1002/wcc.884>
- Leeming J (2023) Searching the web for science: how small mistakes create big problems. *Nature*. <https://doi.org/10.1038/d41586-023-01011-2>
- Leippold M, Vaghefi SA, Stambach D, Muccione V, Bingler JA, et al (2024) Automated fact-checking of climate change claims with large language models. *arXiv:2401.12566*. <https://doi.org/10.48550/arXiv.2401.12566>
- Lelong J, Achache N, Olympie G, Chesneau N, De la Calzada N, et al (2023) ClimateQ&A - a hugging face space by ekimetrics. <https://huggingface.co/spaces/Ekimetrics/climate-question-answering>. Accessed 17 Jul 2024
- Lydari M, El Mourabit Y, El Habouz Y, Fakir M (2022) A performant deep learning model for sentiment analysis of climate change. *Soc Netw Anal Min* 13:8–8. <https://doi.org/10.1007/s13278-022-01014-3>
- Mach KJ, Mastrandrea MD, Freeman PT, Field CB (2017) Unleashing expert judgment in assessment. *Glob Environ Change* 44:1–14. <https://doi.org/10.1016/j.gloenvcha.2017.02.005>
- Majszak M, Jebeile J (2023) Expert judgment in climate science: how it is used and how it can be justified. *Stud Hist Philos Sci* 100:32–38. <https://doi.org/10.1016/j.shpsa.2023.05.005>
- Mastrandrea MD, Mach KJ, Plattner G-K, Edenhofer O, Stocker TF et al (2011) The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. *Clim Change* 108:675–675. <https://doi.org/10.1007/s10584-011-0178-6>
- Mausser W, Klepper G, Rice M, Schmalzbauer BS, Hackmann H et al (2013) Transdisciplinary global change research: the co-creation of knowledge for sustainability. *Curr Opin Environ Sustain* 5:420–431. <https://doi.org/10.1016/j.cosust.2013.07.001>
- OpenAI (2024) Hello-gpt-4o. <https://openai.com/index/hello-gpt-4o/>. Accessed 17 Jul 2024
- Ourbak T, Tubiana L (2017) Changing the game: the Paris Agreement and the role of scientific communities. *Clim Policy* 17:819–824. <https://doi.org/10.1080/14693062.2017.1348331>
- Palutikof JP, Boulter SL, Field CB, Mach KJ, Manning MR et al (2023) Enhancing the review process in global environmental assessments: the case of the IPCC. *Environ Sci Policy* 139:118–129. <https://doi.org/10.1016/j.envsci.2022.10.012>
- Schimanski T, Ni J, Kraus M, Ash E, Leippold M, et al (2024) Towards faithful and robust LLM specialists for evidence-based question-answering. *arXiv:2402.08277*. <https://doi.org/10.48550/arXiv.2402.08277>
- Sietsma AJ, Groenendijk RW, Biesbroek R (2023) Progress on climate action: a multilingual machine learning analysis of the global stocktake. *Clim Change* 176:173–173. <https://doi.org/10.1007/s10584-023-03649-3>
- Sietsma AJ, Ford JD, Minx JC (2024a) The next generation of machine learning for tracking adaptation texts. *Nat Clim Change* 14:31–39. <https://doi.org/10.1038/s41558-023-01890-3>
- Sietsma AJ, Theokritoff E, Biesbroek R, Canosa IV, Thomas A et al (2024b) Machine learning evidence map reveals global differences in adaptation action. *One Earth* 7:280–292. <https://doi.org/10.1016/j.oneear.2023.12.011>
- Stede M, Patz R (2021) The climate change debate and natural language processing. In: *In Proceedings of the 1st Workshop on NLP for Positive Impact*. pp 8–18. <https://doi.org/10.18653/v1/2021.nlp4posimpact-1.2>
- Stocker TF, Plattner G-K (2014) Climate policy: rethink IPCC reports. *Nature* 513:163–165. <https://doi.org/10.1038/513163a>
- Stokel-Walker C, Van Noorden R (2023) The promise and peril of generative AI. *Nature* 614:214–216. <https://doi.org/10.1038/d41586-023-00340-6>
- Thomas A, Theokritoff E, Lesnikowski A, Reckien D, Jagannathan K et al (2021) Global evidence of constraints and limits to human adaptation. *Reg Environ Change* 21:85–85. <https://doi.org/10.1007/s10113-021-01808-9>
- Thulke D, Gao Y, Pelsler P, Brune R, Jalota R, et al (2024) ClimateGPT: towards AI synthesizing interdisciplinary research on climate change. *arXiv:2401.09646*. <https://doi.org/10.48550/arXiv.2401.09646>
- Vaghefi SA, Stambach D, Muccione V, Bingler JA, Ni J et al (2023) ChatClimate: Grounding conversational AI in climate science. *Commun Earth Environ* 4:480–480. <https://doi.org/10.1038/s43247-023-01084-x>
- van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J et al (2021) An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell* 3:125–133. <https://doi.org/10.1038/s42256-020-00287-7>
- Van Dis EAM, Bollen J, Van Rooij R, Zuidema W, Bockting CL et al (2023) ChatGPT: five priorities for research. *Nature* 614:224–226. <https://doi.org/10.1038/d41586-023-00288-7>
- Vardy M, Oppenheimer M, Dubash NK, O'Reilly J, Jamieson D et al (2017) The Intergovernmental Panel on Climate Change: Challenges and Opportunities. *Annu Rev Environ Resour* 42:55–75. <https://doi.org/10.1146/annurev-environ-102016-061053>
- Zommers Z, Marbaix P, Fischlin A, Ibrahim ZZ, Grant S et al (2020) Burning embers: towards more transparent and robust climate-change risk assessments. *Nat Rev Earth Environ* 1:516–529. <https://doi.org/10.1038/s43017-020-0088-0>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Veruska Muccione<sup>1,2</sup>  · Saeid Ashraf Vaghefi<sup>1,3,4</sup> · Julia Bingler<sup>5,6</sup> · Simon K. Allen<sup>1,7</sup> · Mathias Kraus<sup>8</sup> · Glen Gostlow<sup>3</sup> · Tobias Wekhof<sup>3,9</sup> · Chiara Colesanti-Senni<sup>3</sup> · Dominik Stambach<sup>10</sup> · Jingwei Ni<sup>3,9</sup> · Tobias Schimanski<sup>3</sup> · Tingyu Yu<sup>3</sup> · Qian Wang<sup>3</sup> · Christian Huggel<sup>1</sup> · Juerg Luterbacher<sup>4</sup> · Robbert Biesbroek<sup>11</sup> · Markus Leippold<sup>3,12</sup>

✉ Veruska Muccione  
veruska.muccione@geo.uzh.ch

Saeid Ashraf Vaghefi  
saeid.vaghefi@geo.uzh.ch

Julia Bingler  
julia.bingler@smithschool.ox.ac.uk

Simon K. Allen  
simon.allen@geo.uzh.ch

Mathias Kraus  
mathias.kraus@fau.de

Glen Gostlow  
glen.gostlow@df.uzh.ch

Tobias Wekhof  
tobias.wekhof@bf.uzh.ch

Chiara Colesanti-Senni  
chiara.colesantisenni@df.uzh.ch

Dominik Stambach  
dominik.stambach@gess.ethz.ch

Jingwei Ni  
jingwei.ni@gess.ethz.ch

Tobias Schimanski  
tobias.schimanski@df.uzh.ch

Tingyu Yu  
tingyu.yu@df.uzh.ch

Qian Wang  
qian.wang@uzh.ch

Christian Huggel  
christian.huggel@geo.uzh.ch

Juerg Luterbacher  
jluterbacher@wmo.int

Robbert Biesbroek  
robbert.biesbroek@wur.nl

Markus Leippold  
markus.leippold@df.uzh.ch

<sup>1</sup> Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

<sup>2</sup> Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf (Zurich), Switzerland

<sup>3</sup> Department of Finance, University of Zurich, Zurich, Switzerland

<sup>4</sup> World Meteorological Organization (WMO), Geneva, Switzerland

<sup>5</sup> Smith School of Enterprise and the Environment, University of Oxford, Oxford, UK

<sup>6</sup> Council On Economic Policies, Zurich, Switzerland

<sup>7</sup> Institute of Environmental Sciences, University of Geneva, Geneva, Switzerland

<sup>8</sup> Friedrich-Alexander-Universität, Erlangen-Nuremberg, Germany

<sup>9</sup> Center of Economic Research, ETH Zurich, Zurich, Switzerland

<sup>10</sup> Center for Law and Economics, ETH Zurich, Zurich, Switzerland

<sup>11</sup> Wageningen University & Research, Wageningen, The Netherlands

<sup>12</sup> Swiss Finance Institute (SFI), Zurich, Switzerland