# Potential Benefits and Dangers of Using Large Language Models for Advancing Sustainability Science and Communication

Ehsan Nabavi[1], Holger R Maier[2], Saman Razavi[3], Adrian Hindes[4], Mark Howden[5], Will Grant[1], and Sujatha Raman[1]

[1]Australian National Centre for Public Awareness of Science, The Australian National University
[2]Department of Civil and Environmental Engineering, The University of Adelaide
[3]School of Environment and Sustainability, University of Saskatchewan
[4]Fenner School of Environment and Society, Australian National University
[5]Institute for Climate, Energy and Disaster Solutions, Australian National University

March 25, 2024

**Abstract**

Advancements in large language models (LLMs) provide opportunities to accelerate progress towards the attainment of the Sustainable Development Goals (SDGs). Current research largely overlooks the nuanced benefits and dangers LLMs introduce to sustainability research and communication, as well as broader challenges that need to be addressed in the longer term. This paper overcomes these shortcomings by introducing and discussing a framework that highlights how LLMs can benefit knowledge production, mobilization, and communication in the sustainability sciences, as well as any associated dangers.

In addition, it outlines potential long-term challenges that must be acknowledged and addressed to ensure the responsible use of LLMs in advancing sustainability science. A key to the development and use of LLMs for sustainability science is the development of regulatory measures. These measures should be guided by what is needed for expanding sustainability science on the one hand and a holistic view to ensure its responsible use on the other. Failure to reflect and act on this might result in unintended consequences or misuse, making the technology another roadblock to progress towards the SDGs.

# Potential Benefits and Dangers of Using Large Language Models for Advancing Sustainability Science and Communication

**Ehsan Nabavi[1]*, Holger R. Maier[2], Saman Razavi[3], Adrian Hindes[4], Mark Howden[5], Will Grant[1], Sujatha Raman[1]**

[1] *Australian National Centre for Public Awareness of Science, The Australian National University*

[2] *Department of Civil and Environmental Engineering, The University of Adelaide, Australia*

[3] *School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Canada*

[4] *Fenner School of Environment and Society, Australian National University*

[5] *Institute for Climate, Energy and Disaster Solutions, Australian National University*

***Correspondence to**:
Peter Baume Building, The Australian National University, #42A Linnaeus Way, Canberra ACT 2601 Ehsan.Nabavi@anu.edu.au

## Abstract

Advancements in large language models (LLMs) provide opportunities to accelerate progress towards the attainment of the Sustainable Development Goals (SDGs). Current research largely overlooks the nuanced benefits and dangers LLMs introduce to sustainability research and communication, as well as broader challenges that need to be addressed in the longer term. This paper overcomes these shortcomings by introducing and discussing a framework that highlights how LLMs can benefit knowledge production, mobilization, and communication in the sustainability sciences, as well as any associated dangers. In addition, it outlines potential long-term challenges that must be acknowledged and addressed to ensure the responsible use of LLMs in advancing sustainability science. A key to the development and use of LLMs for sustainability science is the development of regulatory measures. These measures should be guided by what is needed for expanding sustainability science on the one hand and a holistic view to ensure its responsible use on the other. Failure to reflect and act on this might result in unintended consequences or misuse, making the technology another roadblock to progress towards the SDGs.

*Submitted to Nature Sustainability*

# MAIN

A recent UN report[1] warns that the world is not on track to reach any of the sustainable development goals (SDGs) by 2030, and scientists may need to approach their work in new ways to accelerate progress [2]. This aligns with the perspective held by some sustainability experts that the dominant research in sustainability science is not adequate for meeting the sustainability agenda and should be expanded [3]. Recent advances in artificial intelligence (AI) to support various aspects of sustainability science [4-7], particularly in large language models (LLMs), could provide new opportunities to attain the SDGs in the requisite timeframes, as evidenced by discussions in the broader scientific community about if and how LLMs could accelerate scientific progress and communication [8-10].

Fields such as Medicine and Chemistry have already embraced LLMs in various applications [11-13]. This includes using them for accelerating drug discovery, optimizing vaccine design, even making mind-reading a tangible possibility [11]. In biotechnology, for example, researchers are actively using LLMs to generate protein sequences with predictable functions across diverse families, demonstrating considerable potential for protein design and engineering [14]. There are even ongoing studies exploring the automation of the entire scientific processes using LLMs, including everything from design of research to planning and execution of lab experiments [15].

In social science, researchers have been inspired to use LLMs to improve agent-based models, simulate human populations taking surveys, and even replacing human participants in social experiments [16-18]. The Impressive ability of LLMs to answer questions has also inspired researchers to use them for creating expert-domain chatbots, in diverse areas such as climate change, finance, physics, material science, and clinical knowledge [19-23].

However, despite their successful application in other domains, the use of LLMs in sustainability science has not yet been explored. This paper aims to fill this gap by providing a holistic framework that helps sustainability researchers to evaluate the potential benefits and dangers of using LLMs to progress and accelerate their work, as well as any associated long-term challenges (Figure 1). This is done via two dimensions of progressing sustainability science, including (1) knowledge production and (2) knowledge mobilization and communication.

Knowledge production committed to diversity, interdisciplinary collaboration, and co-production values is essential for accelerating progress towards the SDGs, as highlighted in the Global Sustainable Development Report [24]. LLMs can provide new opportunities in these domains by synthesizing fragmented evidence in sustainability science and fostering a shared systemic perspective. They can also support experts by developing more interdisciplinary attitudes from diverse areas such as engineering, biology, medicine, social science, law, and more and be used to facilitate consensus-building among stakeholders who often have differing views on problems and solutions, preparing them for socio-political deliberations to expedite actions. In addition, LLMs can create new avenues for co-producing knowledge and conducting intervention experiments in collaboration with local stakeholders, further enhancing the transformative potential of sustainable development initiatives.

Similarly, knowledge mobilization and communication, aimed at making sustainability science universally accessible and mutually beneficial, is also key for accelerating progress towards achieving the SDGs [3]. LLMs can support sustainability experts in these endeavors by providing tools to ensure that the benefits of scientific research are shared equitably. By offering easier ways to access data and enabling cross-lingual translations, LLMs can significantly contribute to narrowing the disparities between the global North and South in their capacities to produce and communicate knowledge.

However, as highlighted In Figure 1, the responsible use of AI in sustainability science requires their potential dangers and long-term challenges to be considered alongside their potential benefits. Such dangers include the introduction of potential biases (technical, language), a lack of competence, legitimacy, awareness, context and/or credibility, the obscuration of flawed reasoning, repressing innovation by reinforcing the established, and the inability to gauge uncertainty. Potential long-term challenges associated with the application of LLMs to advance sustainability science include the production of a massive environmental footprint, a reduction in critical thinking ability, the erosion of scientific integrity, the disruption of progress towards the recognition of the need for nuance and context, contributions towards an AI-driven infodemic, the undermining of human interaction and the risk of expecting increased outputs due to the efficiency savings resulting from the use of LLMs, rather than using this to foster stakeholder engagement and discussion.

The above issues are explored and discussed in the remainder of this paper, which is organized as follows. In the next section, we provide details of how LLMs work, followed by details of the potential benefits and dangers of associated with using LLMs for progressing sustainability science. Finally, we discuss the potential long-term challenges associated with the widespread adoption of LLMs in this field.

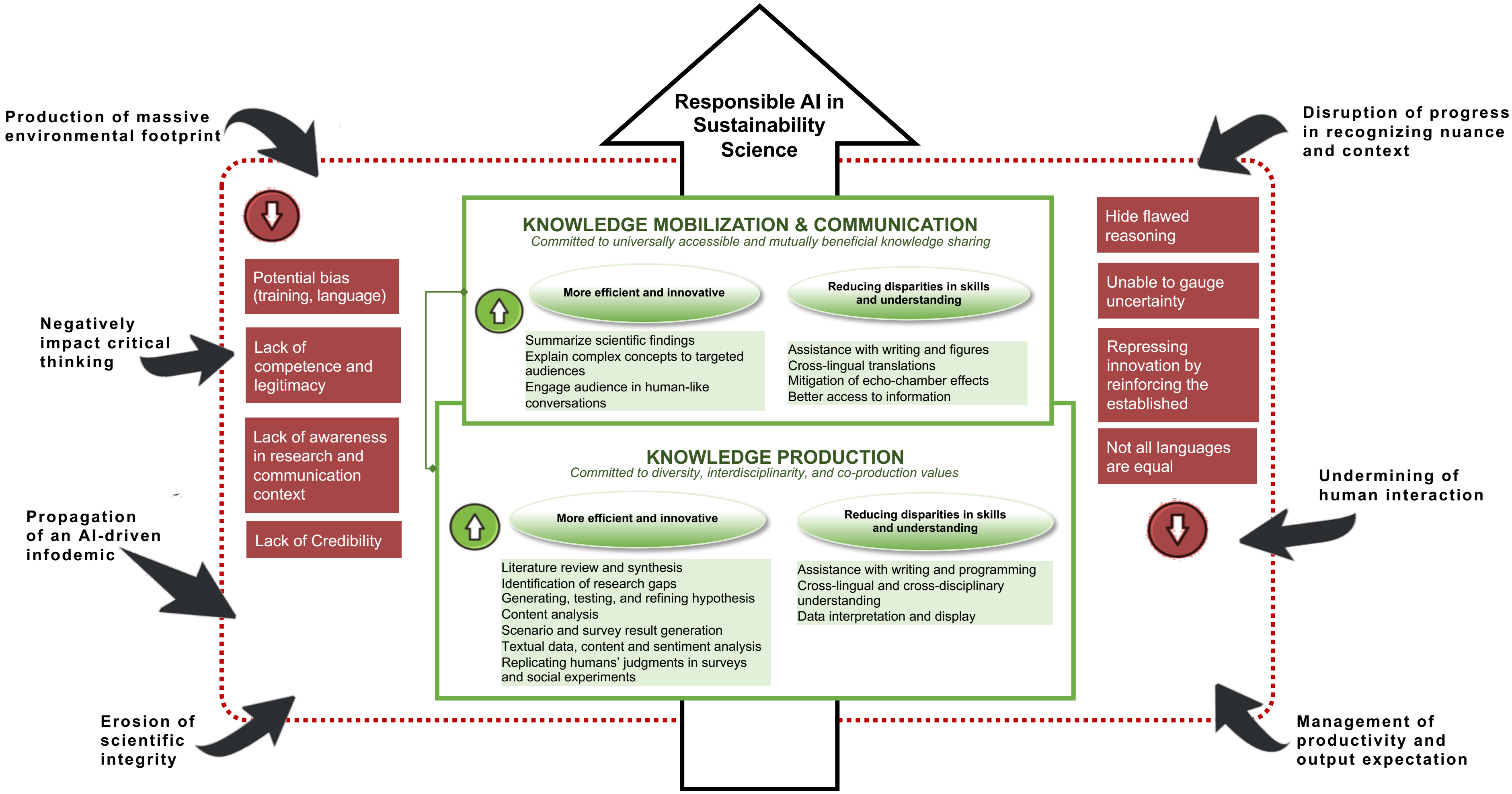


***Figure 1*** *Holistic framework for assessing the impact of LLMs in knowledge production, distribution, and communication—committed to expand sustainability to achieve SDGs. The framework illustrates LLMs' potentials (green boxes), dangers (red boxes), and long-term challenges (black arrows). The goal (vertical arrow) is to ensure responsible AI in sustainability science.*

## WHAT ARE LLMs

The goal of LLMs is to create a model that learns to predict how text fits together in a language sequence. In order to achieve this, the model is trained on an extensive body of unlabelled texts using deep learning, wielding billions of parameters—e.g. 250 billion in ChatGPT. This training process starts with a pre-training phase, in which a neural network model is developed to identify patterns from a corpus of text such as Wikipedia pages, articles, and books. The model uses a 'self-attention mechanism' [25] to capture dependencies between different words in a sentence, and adjust their influence on the output. After the pre-training phase, the model is fine-tuned for specific tasks, for example a chatbot for climate communication. Then, human annotators work with the model to review and rate the model's responses on a scale of quality. This helps the model to learn what constitutes desirable and undesirable responses from a human's perspective. The positive feedback from the annotator acts as a reward in the model, driving it to generate similar responses. This iterative process continues until the model improves its responses. A stronger feedback loop with the human reviewer further aligns the model's responses with guidelines and prespecified ethical principles.

## POTENTIAL BENEFITS

As shown by the green boxes and arrows in Figure 1, LLMs provide a wide range of opportunities to empower sustainability researchers with new tools that have the ability to enhance knowledge production, mobilization, and communication. In both of these areas, these tools fall under the categories of (1) Increasing efficiency and innovation and (2) reducing disparities in skills and understanding, as detailed in the following sub-sections.

### (1) Making sustainability science more efficient and innovative

***LLMs can assist researchers at different stages of research.*** Researchers have already begun to explore and employ LLMs at various stages of the research process, and for different purposes, such as: literature review and synthesis, identifying research gaps and top questions in the field for inspiration see [26], generating, refining, and testing hypothesis, and cross-lingual research understanding through their translation capability.

The landscape in this domain is evolving rapidly. There are online tools designed specifically for these tasks, with some having already shown potential in aiding meta-analysis and data extraction. Elicit (https://elicit.com/), for example, has been trained on a vast array of scientific texts to streamline literature reviews. It is particularly fine-tuned for answering empirical research with interventions and questions such as : What are the effects of '*climate change*' on '*coastal ecosystems*'? or Does '*urbanization*' impact '*water quality in rivers*'?

With the recent releases of third-party plugins for GPT-4, at the time of writing, LLMs can now directly access the web and research archives to assist in the initial steps of literature reviews. The development of research-oriented plugins such as ScholarAI (direct access to peer-reviewed articles) and Wolfram (real-time computation and data) suggests more curated LLMs for scientific research may be on the horizon. Open-source versions of LLMs, such as OpenLLaMA and Pythia, have also begun proliferating, opening the door to wider adoption and use on local computers rather than the cloud.

Given these capabilities, LLMs present a unique opportunity to progress sustainability science by making it easier to access and synthesize information from a wide range of disciplines (processing extensive bodies of literature, extracting mainstream arguments, minority views, and diverse perspectives on sustainability problems) and possibly facilitating interdisciplinary collaboration and co-production. This capability enables researchers to gain a better understanding of the varied interpretations of SDG-related issues, both locally and globally, while promoting transparency in science production. Through synthesizing dimensions and perspectives from diverse sources, LLMs contribute to creating more interdisciplinary narratives. They can also help researchers to form connections between disparate concepts and ideas from various disciplines intersecting in sustainability science, aiding them in discovering new insights and interdisciplinary connections. On a broader scale, the use of LLMs has the potential to accelerate ongoing efforts to synthesize fragmented evidence in the field of sustainability, thus, facilitating policy breakthroughs similar to the 2015 Paris Climate Agreement.

***LLMs streamline data analysis.*** LLMs have also proven useful in qualitative research, particularly for content analysis and text-based data. Research shows that LLMs perform reasonably well in qualitative coding when compared to hand-coded data-bases from various fields such as sociology, psychology, and linguistics [17]. They can be used to analyze social media posts, news articles, and government reports on environmental-related topics, synthesizing public and policy discourse. Leveraging this capability, LLMs have already been deployed to streamline the analysis of extensive reports containing vast amounts of information. In one example, an LLM has been developed to automate the analysis of corporate sustainability reports against the Task Force for Climate-Related Financial Disclosures (TCFD) recommendations—an effort that is feasible for only a select few entities globally due to associated resource constraints [27]. Similarly, Toetzke, et al. [28] suggest using LLMs for solution scanning, allowing policymakers to monitor ever-growing data on climate technology innovation in policy documents, social media data, and company reports. This could support researchers to accelerate evidence synthesis in climate and environmental reporting on a range of issues from social aspects to governance, positioning them as artificial 'policy advisers' [29] in sustainability science.

Beyond content analysis, LLMs can also contribute to interpreting extensive geospatial data generated by climate and hydrologic models. The efficiency and versatility of this capability can be improved further by incorporating third-party plugins.

***LLMs inspire innovative research approaches.*** LLMs offer the potential for adopting more sophisticated and creative research approaches. For example, LLMs have been found capable of replicating humans, and their moral judgments, in social experiments. More specifically, Dillion, et al. [30] showed that there is a 0.95 correlation between average human judgments and GPT-3.5 across 464 moral scenarios sourced from peer-reviewed papers. This allows researchers to develop human-like samples by creating imaginary participants, giving them certain qualities and features, placing them in different scenarios, and studying how they behave. This capability has empirically been demonstrated in different social studies from vote prediction [31] to economic games [32].

Researchers can use this 'silicon sampling' to simulate a diverse population of participants in their studies, addressing issues such as energy and water consumption, sustainable food choices, and transportation preferences. This capability offers new research infrastructure, where science can be produced through interaction with non-human participants, or "homo silicus" [32]. This allows researchers to conduct surveys and experiments rapidly, reliably, without cost, and at scale, potentially marking the beginning of a synthetic-social science where researchers no longer have to worry about human recruitment, incentives, and applying for ethics approval for their social research.

This capability of LLMs offers sustainability researchers new research opportunities. For example, it allows them to generate scenarios or narratives based on potential pathways or interventions, in domains such as water and climate, and then test them with local stakeholders. This could offer researchers innovative approaches to engage in co-producing knowledge, hands-on testing, and developing novel methods for collecting and testing place-based pathways for transformation. This shift should empower researchers to transition from individual research modes to cooperative approaches [3].

***LLMs make science communication more efficient***. LLMs can streamline the process of generating content ideas, to continuously improve their translational ability to summarize scientific findings and to explain complex concepts and subjects to ever more precisely targeted audiences, as well as assisting science communicators to write op-eds, social media posts, and design social campaigns. This facilitates adjustment of linguistic complexity, terminology and scientific findings according to project objectives and the needs of target groups, such as different ages or educational levels. LLMs can also be customized to serve as an artificial dialogic agent, engaging users in seemingly real conversations where various aspects of a sustainability problem can be explored until they are completely satisfied, which is similar to working with an interactive Wikipedia, where users can delve deeper just by asking questions and following up on aspects of the answer they may not have fully understood. This could facilitate dialogical science communication on a large scale, presenting an opportunity to broaden and democratize sustainability conversations, which was previously limited to small groups of science enthusiasts [8].

## (2) Reducing disparities in skills and understanding

***LLMs help with writing and programming***. LLMs are already seeing adoption as writing assist tools. GPT-3.5 and 4 can kickstart the outlining process for a paper or help with becoming unstuck in the writing process, allowing researchers to focus more on research experiments and developing new ideas. This capability is particularly helpful for non-native English speakers in writing papers and translating research across languages, which ultimately could increase the representation of researchers from non-English speaking countries. In fields such as sustainability science, this can mean more localized sustainability issues and perspectives are able to receive recognition on a global scale. For English researchers, LLMs can be useful in helping to identify the weaknesses of arguments and counter-arguments, or reflect on how readers might interpret them.

LLMs can also assist researchers who have limited computer programming skills to write, debug and annotate computer code, and translate code from one programming language to another. Such tools include GPT-4 and programming-specific tools such as Github Copilot that are designed to assist with more specialized tasks such as programming workflows [33]. This could potentially contribute to democratizing science by allowing researchers who have less expertise in coding or lack the resources to hire a full-time programmer to undertake research that requires programming. Although unlikely to write full models by themselves yet, smaller problems such as implementing certain algorithms, utilizing APIs and making sense of obtuse documentation can be made more tractable with copilot-style LLMs.

The above capabilities of LLMs can therefore contribute to progress in sustainability science by increasing accessibility and reducing inequity, especially in the basic skills that are crucial in today's research environment, such as the ability to code and to write in English. This will be beneficial for addressing the divide between the Global North and South and between disciplines, including disparities in data access and research capacity between high-, low-, and middle-income countries. According to Messerli, et al. [3], achieving the SDGs is unlikely without addressing this inequality. Consequently, LLMs can be a useful tool to support researchers globally in their mission to promote universally accessible and mutually beneficial sustainability science.

***LLMs help mitigate echo chamber effect***. The design of LLMs allows science communicators to address some aspects of the pervasive challenge of mis/disinformation. By effectively gathering and synthesizing information, these models may serve to reduce over-exposure to fringe and incorrect claims, by making a statistical mean for topical knowledge. This has the potential to produce a 'social consensus model' that represents generalized social agreement on a topic. For example, when an individual asks an LLM about the evidence base for Anthropogenic Climate Change, in theory the output should be a social consensus model of the average human understanding of the genesis of climate change, which has already been fine-tuned by the model owner and regulators. Consequently, LLMs have the potential to overcome issues such as the overrepresentation of niche or contrarian notions in the marketplace of ideas, and instead be a hub for understanding what the general consensus on a topic is, accounting for a diverse array of inputs.

More specifically, LLMs can be trained on a particular knowledge domain. For example, BloombergGPT is a 50 billion parameter model that is trained on a wide range of financial data [21]. Another example is ChatClimate, which is an LLM that is trained on the latest IPCC Report, with the vision to make climate risk understandable and climate information more accessible to the broader community [20].

While these instances primarily represent the broadcast or didactic aspects of science communication, focusing less on dialogue, deliberation, or participation, it is important to recognize the valuable opportunities LLMs provide in this context. Specifically, sustainability science can benefit significantly by using these tools to foster consensus among stakeholders on scientific findings, and to minimize disagreements about problems and solutions (DeFries & Nagendra, 2017). In addition, as discussed, LLMs can also support experts in creating platforms that synthesize disciplinary and interdisciplinary evidence on key trade-offs and co-benefits across various contexts and scales. This has the potential to assist with responding to the call to accelerate gathering, aggregating, and evaluating information on both successful and failed interventions, from diverse knowledge systems arising from lay, practical, and indigenous sources [3].

# POTENTIAL DANGERS

As shown in the small red boxes in Figure 1, LLMs can also pose potential dangers to the progress of sustainability science if researchers fail to exercise caution when integrating them into the processes of knowledge production, distribution, and communication. These dangers underscore the need for deliberate consideration and thoughtful integration of LLMs into research and communication practices to ensure their responsible use, as detailed below.

***LLMs hide flawed reasoning***. There is general consensus that reasoning is the Achilles heel of all LLMs. Yet, models such as ChatGPT are rapidly improving, beyond cutting and pasting material from their training phases, and repeating them like a parrot. GPT-4 has been shown to be able to learn meaning even if it has only been trained to predict the next token in a text [34]. This means that LLMs can no longer be referred to as 'stochastic parrots' that only learn from the statistics of syntax [35]. Having said that, at present, GPT-4 is still incapable of reasoning [36], and this Is likely to make it more challenging for users to identify LLMs' potential faulty logic or incorrect responses. This is of significant concern as these models become better at generating text that seems more and more legitimate, particularly when they are trained on a certain subject matter such as water or climate.

The above concerns underscore the need for researchers to exercise caution and conduct their due diligence to ensure the accuracy and credibility of the information provided by LLMs. This task becomes particularly complex when dealing with emerging, less-understood sustainability problems viewed from multiple perspectives. A case in point is the impact of permafrost thaw on the hydrology of northern rivers and its subsequent effects on water supply to neighboring communities, an area where outcomes are varied and largely yet to be fully understood.

Researchers should also maintain a humble and rigorous approach, following the scientific method, if they use LLMs in their work. Non-positivist epistemologies, which, being more creative and less descriptive, are less likely to be successfully emulated by LLMs. This raises another concern: what if the original text used for training is itself based on faulty reasoning? Non-experts in a subject matter could easily mistake incorrect information for accurate knowledge. These issues could be amplified in fields such as hydrology and environmental science, where existing literature is not as extensive as in other research domains.

***LLMs lack awareness of research context.*** As discussed above, LLMs are limited in perceiving, reasoning, and planning based on parameters grounded in reality. In other words, they are unable to meaningfully engage with ontological and epistemological questions that shape scientists' understanding of reality and inform their research inquiries and methods. As a result, LLMs can easily misinterpret research findings as they are not able to capture the implicit value and judgments in scientific writings [37], as well as the boundary judgments in modeling and simulations [38]. In addition, they are unable to capture the effectiveness of results stemming from oversimplified or excessively complex models, as well as potential ethical concerns. This means that LLMs simply cannot capture the limitations and nuances of research that are obvious to scientists and their epistemic communities. For example, a water crisis is a multifaceted challenge driven by various environmental, social, and geo-political factors. LLMs may not be able to grasp the nuanced context in which related research questions are posed and specific solutions are sought.

***LLMs lack competency and accountability***. LLMs might show a high level of performance in undertaking tasks in various domains (as discussed in the previous section), but this does not mean they are competent at it. When we see someone perform well in analyzing datasets, we tend to assume they are competent in the field of data science. However, we cannot generalize this idea to LLMs, which may show acceptable performance with plug-ins such as the Code Interpreter but lack competence as a quality of expertise. For this very reason, they lack accountability for their responses and actions, which may raise ethical concerns. Who should bear the responsibility for any unintended consequences that may be caused by their use? What about the reproducibility and the scientific method when LLMs, by design, blackbox the underlying mechanisms that generate the output? Does this undermine the inference drawn from LLMs?

***LLMs are biased***. While we might expect LLMs to provide a balanced representation of diverse opinions (i.e. via their social consensus mode explained earlier), they are unable to reflect popular opinion. This is because LLMs have built-in biases shaped by: (1) the training data and (2) the fine-tuning process moderated by annotators and companies. For example, in research conducted on 60 US demographic groups over topics ranging from abortion to automation, LLMs exhibited misalignment with public opinion and those of various demographic groups [39]. When the model was trained on internet data alone, the responses were biased toward less educated, lower income, or conservative points of view. In contrast, newer models that were fine-tuned through human feedback had left-leaning tendencies, biased towards more liberal, higher educated, and higher income audiences.

The above issue highlights that the fine-tuning method could act as a double-edge sword, potentially introducing cultural and political biases leading to a spread of disinformation on an unprecedented scale. This limitation of LLMs for science communication, however, will continue to exist until a comprehensive system of human knowledge is developed through extensive crowdsourcing fine-tuned by millions of expert-volunteers, similar to the way Wikipedia currently operates.

***LLMs are unable to gauge uncertainty***. Positivist methods that use quantitative data, typical in sustainability science, usually offer uncertainty estimates alongside their results. Characterizing, reducing, and communicating uncertainty are all vital steps in tackling sustainability challenges. However, the conventional positivist perspectives on uncertainty, rooted in probability theory and beyond, cannot be directly applied to the outputs of LLMs. This prompts the6uestionn: How can a user discern what level of confidence to place in a given response? Moreover, LLMs fundamentally cannot disclose to the user the specific body of text that was used to answer a question, nor can they comment on the credibility of those sources, although there are attempts to address this issue [40,41].

***LLMs reinforces established research and repress innovation***. LLMs, by their nature, are trained on existing bodies of text, which inherently encapsulate the mainstream views and ideas of the time they were written. While reinforcing existing knowledge is generally valued, there is a danger that this could inadvertently hinder research innovation. The underlying data used in LLMs are largely limited to open access text from the internet, which is a massively biased data set, and privileges particular types of knowledge from particular types of people. There is also the risk of available narratives converging around dominant ways of understanding the world and the environment (e.g. the commodification view over the conservation view of nature). The training data used by LLMs can therefore perpetuate and deepen existing biases regarding complex environmental issues such as environmental racism, climate change, global environmental change, biodiversity loss, and pollution [42].

Consequently, LLMs may perpetuate these dominant perspectives in their responses, which, over time, can create a positive feedback loop. This can inadvertently cement the status quo, reinforcing mainstream research ideas and allowing them to grow further. This trend can stifle the exploration of alternative viewpoints or innovative solutions, which could potentially hinder progress in addressing complex sustainability challenges.

***LLMs lack credibility and legitimacy.*** Complex problems in sustainability science demand intensive interactions among stakeholders and researchers to co-produce knowledge that is salient, credible, and legitimate. LLMs may be able to increase salience through more timely and comprehensive analysis of data [43]. However, it is an open question at this stage as to whether LLMs could someday provide credible or legitimate information given their foundational limitations, including: bias and misrepresentation, lack of traceability of output, absence of representation of the underlying physical processes, and the absence of information about uncertainties or sensitivity analysis. Moreover, resolution of such complex issues often requires establishment of new, highly contextual solutions (i.e. 'extrapolation'), which is likely to be at odds with the inherent 'interpolative' nature of LLMs.

***Not all languages are equal in LLMs***. The ongoing discussion about LLMs predominantly centers on user experiences in English. For example, ChatGPT's responses in non-English languages are of lesser quality than those in English, with outputs often containing inaccurate or nonsensical sentences. Unequal access to LLMs among societies globally also has the potential to exacerbate existing digital differences within and between communities in favor of those who have access to the required infrastructure and environmental information.

# FUTURE CHALLENGES

The black arrows in Figure 1 represent future challenges that must be acknowledged and addressed to ensure the responsible use of LLMs in advancing sustainability science. This requires researchers to reflect and take action to minimize associated risks, while harnessing the potential benefits of LLMs. Overlooking these considerations may make LLMs an inhibitor, rather than an enabler, of efforts to accelerate the SDG agenda.

***LLMs have a massive environmental footprint.*** LLMs are emerging as a growing techno-solutionist paradigm within which AI is often touted as a crucial tool for tackling climate change [44]. However, the growth of AI itself is already contributing to carbon emissions [45], thereby potentially exacerbating the climate crisis.  This is because the training and inference processes in LLMs require significant amounts of energy and water for both computational power and data center cooling— which underscores the need for energy-efficient algorithms [46]. For example, training GPT-3 models consumes about 1,287 MWh and generates 552 tons of carbon dioxide equivalent [47]. These numbers, however, vary based on the size and geographic location of the model, as well as datacenter infrastructure. For instance, the above energy consumption values could triple if training is conducted in data centers in Asia [48].

Environmental impacts also occur during the use of LLMs. For example, once users start interreacting with LLMs, the inference engine kicks in, which consumes around 500ml of water, even for a simple chat [48]. As a result, Microsoft [49] has already reported a 34% surge in its annual water consumption, which is primarily attributed to its investments in AI and partnership with OpenAI.

Thus, we anticipate that as the use of LLMs continues to grow, the ecological footprints of these models will increase, imposing a much heavier toll on the planet compared to their predecessors. This issue becomes more critical when countries with different efficiency standards may decide to develop their own models and keep their own sovereignty over them. This underscores the profound implications of LLMs on sustainability and the ethical dilemmas and trade-offs associated with their use [50].

***LLMs raise concerns about scientific integrity.*** Reproducibility is the cornerstone of any scientific endeavor, especially in terms of fostering trust in research outcomes. This is particularly important for sustainability science, where solutions are emerging at the intersection of environment, society, and policy and demand thorough validation. To achieve this, researchers using LLMs need to be transparent about the structure of the model, training datasets, inputs, and initial conditions. However, the black-box nature of these models and the reluctance of corporations controlling the pre-training phase to share information reduces the transparency of the research process, running counter to the principles of open science.

Where advances In LLMs reach the point where they become 'ResearchBots' capable of mimicking most research activities [51], this lack of transparency and reproducibility raises serious concerns about scientific integrity, particularly within the prevailing culture of 'publish or perish' in academia. Emerging responses to address these concerns include the implementation of new rules for LLM usage in academic journals and conferences, as well as a deeper reflection on science in the age of LLMs. This discussion certainly needs broader engagement and action within all scientific communities.

***LLMs could undermine the value of human interaction.*** LLMs can replace human labor at various stages of the research process, and as discussed earlier, allow researchers to rely on synthetic participants for social experiments to study people's behavior, emotions, and motivation. This might undermine the value of human interaction, which is central to the social study of sustainability challenges. One Silicon Valley company, which provides LLM-based services for social research, explicitly states in its mission that researching with synthetic participants/users blurs the long-lasting distinction between qualitative and quantitative methods, stating, "running 10 interviews to gauge [a feature in the society] or deploying 5,000 surveys can now demand roughly the same amount of effort." [52]

***LLMs could disrupt progress in recognizing nuance and context.*** The complexity of global challenges, including sustainability problems, demands nuanced and contextualized expert judgment for public policy decisions [53]. Simplistic appeals to 'trust the science' tend to backfire as the evidence base is often contested, incomplete, ambiguous and liable to be interpreted in different ways. Moreover, the seemingly universal standards derived from experimental science also need to be reconciled with local realities [54].

The task of bringing science to bear on sustainability challenges is therefore more akin to an art of generating 'serviceable truths' [55], that is appropriate to the complexity of problems and potential solutions. This requires diverse inputs from across different disciplines (natural sciences and social sciences) that need to be assessed and put in context of the language, culture and value commitments at play in specific policy domains [56]. However, the rise of LLMs threatens to disrupt the progress that has been made in recognizing the need for nuance and attention to context in this regard. This is because LLMs are not able to provide the type of public reasoning-for-policy necessitated by the complexity of sustainability challenges as they rely upon statistical integration of pre-existing inputs, rather than negotiation and dialogue.

***LLMs could cause an AI-driven infodemic.*** LLMs' power in generating high volumes of authoritative-looking text allows them to be used for dissemination of persuasive misinformation and destructive disinformation. Misinformation results from an unintentional infusing of bias in the training data, the algorithm, and the output, while disinformation is about doing the above deliberately.

LLMs can excel at creating misinformation because they can provide additional details and mimic personal tones quite effectively, often surpassing human capabilities [57]. As a result, conventional assessment guidelines, which rely on criteria such as evidence credibility, source transparency, and acknowledgment of limitations, might not be effective at identifying misinformation.

LLMs can also become a tool for generating disinformation for adversarial purposes. This is because they are generally vulnerable to a threat called 'model spinning' [58], where the output can be manipulated to support a specific viewpoint or hypothesis, when the prompt includes specific words. LLMs, thus, have the potential to become an adversarial machinery for generating disinformation at scale with a deceptive level of detail and accuracy.

All of this distorted information might then fuel a flood of untrustworthy science in society, particularly around contentious topics such as water, climate, and energy issues in crisis times. Furthermore, when combined with other modalities such as image, audio, and video, the production of misinformation related to controversial sustainability issues could both quantitatively and qualitatively be intensified. LLMs could additionally be used by malicious actors to analyze data on a contentious environmental issue, such as climate change or water security, and derive false but apparently authoritative results with motivated reasoning. This result would be a proliferation of synthetic lies at an overwhelming scale, undermining basic human senses in discerning real from fake, wrong from right—severely impacting society's trust in science, scientists, and their institutions.

***LLMs may negatively impact critical thinking***. The ability to challenge assumptions and apply reasoning is vital for making informed decisions and solving complex problems. Many scientific breakthroughs have occurred by turning consensus on ideas completely on its head, what Thomas Kuhn calls 'paradigm shift'. Do LLMs allow us to accommodate new ideas given their inference engine is built on an 'attention mechanism' designed to generate insights?

The increasingly dominant 'attention economy' has led to a very deliberate saturation of society by oversimplified ideas. Combined with social compact being diverted from pro-social to hyper-individualist norms, the use of knowledge 'crutches' is all around us in the contemporary world, even in the absence of LLMs, and as such, deep knowledge and critical thinking are at risk of significant decline. LLMs may rapidly accelerate this process. Instead of doing the hard yards in the literature to deeply understand a topic, researchers or policymakers could instead ask an LLM to produce a very convincing output. Afterall, if society rewards shallow understandings and pursuits, why should the researcher or policymaker bother to actually understand anything?

***LLMs could become catalysts that ask for more.*** There is a deeper concern that we will observe Jevons' Paradox and the myth of technological liberation as a result of the introduction of LLMs. Just as we are in a crisis of overconsumption and overproduction from the natural world, the same mindset of MORE pervades how we relate to knowledge and content. LLMs should free up sustainability researchers to spend less time in the office and behind a computer, and more time meeting with stakeholders outside of the academy. The same applies to policymakers. If we need less screen time because we can achieve the same outputs more efficiently with the help of LLMs, then we need to make sure we do not simply demand more outputs. The saved resources (energy, time, etc.) should be treated as an opportunity for actively engaging with people in communities, experiencing environmental change, and innovating on environmental challenges.

## RESPONSIBILITY MUST BE AT THE CORE

Scientists entered the 21st century with a new social contract [59] to support science in the face of unprecedented environmental and social changes. At the core were the three challenges of new fundamental research, faster knowledge transfer to policymakers, and better public communication.

With AI reshaping the field of sustainability science (Nabavi et al., 2019) and demonstrating significant potential in supporting SDGs (Vinuesa et al., 2020), LLMs can support scientists in fulfilling their roles within this social contract. This is particularly timely, given the increasing call to "change gears" in research practices to enable the SDGs to be achieved in a timely manner [60].

So far, we have explored the potential benefits, dangers, and future challenges associated with integrating LLMs into sustainability science. Yet many deep questions remain that require attention from research communities involved in advancing the SDG agenda (see Table 1). For example: Is using LLMs in research a form of 'outsourcing' science, and does this violate the social contract for public funding of science? How can we involve the public to ensure that researchers understand and meet public expectations and needs as they rely on LLMs in their research activities? How will these developments impact collaborations between different fields, both within a country and internationally? And many more.

The key principle that can guide us in addressing these questions is: responsible integration of AI systems in sustainability science. In Figure 1, this is shown by the vertical arrow, extending across both knowledge production and communication, which should guide us through the journey. Following that logic, it is essential to prioritize 'responsibility' in the design, development, and deployment of LLMs in research and science communication. The good news is that LLMs are still in the early phases of their evolution to become embedded in our choices, the same way many software we rely on in our work, already are. Thus, it is timely for communities of researchers and practitioners focused on advancing the SDG agenda to develop a systemic perspective in understanding AI systems [61], and to establish standards and robust regulatory mechanisms to ensure responsible use in their work. New policies are needed to mitigate their risks and address future challenges, as discussed briefly in this paper. And more granular analyses are needed to capture the complexity of the risk–benefit trade-offs associated with using them, which will allow us to leverage their potential responsibly.

***Table 1*** *Questions that require attention from research communities involved in advancing the SDGs agenda*

| | **QUESTIONS** |
|---|---|
| **OBJECTIVES AND SAFEGUARDS** | - What specific research objectives can be addressed using LLMs in sustainability science? What are the areas in which LLMs should be cautiously used? What are the safeguards that need to be in place for that?<br>- How can research findings be effectively communicated using LLMs? What are the safeguards that need to be in place for that?<br>- How can sustainability experts best assess, and address LLMs' known risks to avoid unreliable outputs? |
| **CONTEXTUAL UNDERSTANDING AND BIASES** | - How much contextual understanding is required when using LLMs?<br>- How can biases in the training data, methods, and resulting outputs be identified and mitigated?<br>- What steps can be taken to ensure that LLMs used for communication purposes do not perpetuate biases or misinformation? |

| | |
|---|---|
| **AUTHORSHIP, PLAGIARISM, AND INTELLECTUAL PROPERTY** | - Does using LLMs mean we need to revisit the very notion of authorship, plagiarism, and intellectual property? What are new definitions and how can they be applied in scientific work? What mechanisms are needed to realize this change? |
| **EXPLAINABILITY AND TRANSPARENCY** | - How can the blackbox nature of LLMs be integrated into research activities without undermining scientific rigor and credibility?<br>- How can the key sources in the training process be traced back?<br>- How does the lack of explainability in LLMs pose challenges for identifying errors, biases, and ensuring transparency in sustainability research? |
| **OVERRELIANCE AND CRITICAL THINKING** | - What potential consequences might arise from overreliance on LLMs for doing research and communication around SDG-related topics?<br>- How can sustainability experts strike a balance between using LLMs and maintaining critical thinking and scientific rigor? |
| **REPRODUCIBILITY** | - What steps can be taken to ensure the reproducibility of research findings and the models' performance? |
| **LIMITATIONS AND UNCERTAINTIES** | - How can LLMs effectively convey the limitations and uncertainties associated with their use?<br>- How can LLMs be leveraged to extract insights about anomalies when studying under-researched phenomena, while maintaining reliability. |
| **COLLABORATION** | - What are the key domains of interdisciplinary collaboration to advance sustainability science in the era of LLMs? What measures and strategies can be employed in the shorter and longer-term?<br>- How can science communicators collaborate with researchers and experts to ensure accurate and reliable communication of complex topics through LLMs? How can they contribute to the responsible and sustainable use of LLMs in their communication work?<br>- How can social and environmental concerns be integrated into the design and use of LLMs?<br>- What educational programs and training initiatives are needed to equip researchers with the skills and knowledge required for responsible sustainability science in the context of LLMs?<br>- In what ways can public engagement be integrated into interdisciplinary efforts?<br>- What mechanisms and platforms can be developed to foster international cooperation and data sharing among teams working on sustainability science with LLMs? |

## ACKNOWLEDGMENTS

We extend our appreciation to Rebecca Colvin and Glenn Althor for their support during the initial drafting stages of this paper.

## REFERENCES

1 UN. Global Sustainable Development Report 2023: Times of crisis, times of change: Science for accelerating transformations to sustainable development. (United Nations, New York, 2023).

2 Malekpour, S. *et al.* What scientists need to do to accelerate progress on the SDGs. *Nature* **621**, 250-254 (2023).

3 Messerli, P. *et al.* Expansion of sustainability science needed for the SDGs. *Nature sustainability* **2**, 892-894 (2019).

4 Nabavi, E., Daniell, K. A., Williams, E. T. & Bentley, C. M. AI for sustainability: a changing landscape. *Artificial intelligence—for better or worse. Future Leaders*, 157-176 (2019).

5 Dauvergne, P. *AI in the Wild: Sustainability in the Age of Artificial Intelligence*. (MIT Press, 2020).

6 Maier, H. R. *et al.* Exploding the myths: An introduction to artificial neural networks for prediction and forecasting. *Environmental modelling & software*, 105776 (2023).

7 Razavi, S. *et al.* Coevolution of machine learning and process-based modelling to revolutionize Earth and environmental sciences: A perspective. *Hydrological Processes* **36**, e14596 (2022).

8 Schäfer, M. S. The Notorious GPT: science communication in the age of artificial intelligence. *Journal of Science Communication* **22**, Y02 (2023).

9 Wang, H. *et al.* Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47-60 (2023).

10 Van Noorden, R. & Perkel, J. M. AI and science: what 1600 researchers think. *Nature* (2023). https://doi.org:https://doi.org/10.1038/d41586-023-02980-0

11 Tang, J., LeBel, A., Jain, S. & Huth, A. G. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, 1-9 (2023).

12 Pan, J. Large language model for molecular chemistry. *Nature Computational Science* **3**, 5-5 (2023). https://doi.org:10.1038/s43588-023-00399-1

13 Li, S. *et al.* CodonBERT: Large Language Models for mRNA Design and Optimization. *bioRxiv*, 2023.2009. 2009.556981 (2023).

14 Madani, A. *et al.* Large language models generate functional protein sequences across diverse families. *Nature Biotechnology* **41**, 1099-1106 (2023). https://doi.org:10.1038/s41587-022-01618-2

15 Boiko, D. A., MacKnight, R. & Gomes, G. Emergent autonomous scientific research capabilities of large language models. *arXiv preprint arXiv:2304.05332* (2023).

16 Aher, G., Arriaga, R. I. & Kalai, A. T. Using large language models to simulate multiple humans. *arXiv preprint arXiv:2208.10264* (2022).

17 Ziems, C. *et al.* Can Large Language Models Transform Computational Social Science? *arXiv preprint arXiv:2305.03514* (2023).

18 Bail, C. A. Can Generative AI Improve Social Science? (2023). https://doi.org:10.31235/osf.io/rwtzs

19 Caruccio, L. *et al.* Can ChatGPT provide intelligent diagnoses? A comparative study between predictive models and ChatGPT to define a new medical diagnostic bot. *Expert Systems with Applications* **235**, 121186 (2024).

20 Vaghefi, S. A. *et al.* Chatclimate: Grounding conversational AI in climate science. (2023).

21 Wu, S. *et al.* Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564* (2023).

22 Singhal, K. *et al.* Large language models encode clinical knowledge. *Nature* **620**, 172-180 (2023). https://doi.org:10.1038/s41586-023-06291-2

23 Xie, T. *et al.* DARWIN Series: Domain Specific Large Language Models for Natural Science. *arXiv preprint arXiv:2308.13565* (2023).

24 GSDR. Independent Group of Scientists appointed by the Secretary-General. Global sustainable development report 2019: the future is now–science for achieving sustainable development. (UN, New York, 2019).

25 Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).

26 Lahat, A. *et al.* Evaluating the use of large language model in identifying top research questions in gastroenterology. *Scientific Reports* **13**, 4164 (2023). https://doi.org:10.1038/s41598-023-31412-2

27 Ni, J. *et al.* Paradigm Shift in Sustainability Disclosure Analysis: Empowering Stakeholders with CHATREPORT, a Language Model-Based Tool. *arXiv preprint arXiv:2306.15518* (2023).

28 Toetzke, M., Probst, B. & Feuerriegel, S. Leveraging large language models to monitor climate technology innovation. *Environmental Research Letters* **18**, 091004 (2023).

29 Tyler, C. *et al.* AI tools as science policy advisers? The potential and the pitfalls. *Nature* (2023). https://doi.org:https://doi.org/10.1038/d41586-023-02999-3

30 Dillion, D., Tandon, N., Gu, Y. & Gray, K. Can AI language models replace human participants? *Trends in Cognitive Sciences* (2023).

31 Argyle, L. P. *et al.* Out of one, many: Using language models to simulate human samples. *Political Analysis* **31**, 337-351 (2023).

32 Horton, J. J. Large language models as simulated economic agents: What can we learn from homo silicus? , (National Bureau of Economic Research, 2023).

33 Zhang, B., Liang, P., Zhou, X., Ahmad, A. & Waseem, M. Practices and Challenges of Using GitHub Copilot: An Empirical Study. *arXiv preprint arXiv:2303.08733* (2023).

34 Jin, C. & Rinard, M. Evidence of Meaning in Language Models Trained on Programs. *arXiv preprint arXiv:2305.11169* (2023).

35 Arkoudas, K. ChatGPT is no stochastic parrot. But it also claims that 1 is greater than 1. *Philosophy & Technology* **36**, 54 (2023).

36 Arkoudas, K. GPT-4 Can't Reason. *arXiv preprint arXiv:2308.03762* (2023).

37 Kasirzadeh, A. & Gabriel, I. In conversation with Artificial Intelligence: aligning language models with human values. *Philosophy & Technology* **36**, 1-24 (2023).

38 Nabavi, E., Daniell, K. A. & Najafi, H. Boundary matters: the potential of system dynamics to support sustainability? *Journal of Cleaner Production* **140**, 312-323 (2017).

39 Santurkar, S. *et al.* Whose opinions do language models reflect? *arXiv preprint arXiv:2303.17548* (2023).

40 Gou, Z. *et al.* Critic: Large language models can self-correct with tool-interactive critiquing. *arXiv preprint arXiv:2305.11738* (2023).

41 Tian, K. *et al.* Just ask for calibration: Strategies for eliciting calibrated confidence scores from language models fine-tuned with human feedback. *arXiv preprint arXiv:2305.14975* (2023).

42 Rillig, M. C., Ågerstrand, M., Bi, M., Gould, K. A. & Sauerland, U. Risks and benefits of large language models for the environment. *Environmental Science & Technology* **57**, 3464-3466 (2023).

43 Cheong, S. M., Sankaran, K. & Bastani, H. Artificial intelligence for climate change adaptation. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **12**, e1459 (2022).

44 Rolnick, D. *et al.* Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)* **55**, 1-96 (2022).

45 Zhong, J., Zhong, Y., Han, M., Yang, T. & Zhang, Q. The impact of AI on carbon emissions: evidence from 66 countries. *Applied Economics*, 1-15 (2023).

46 An, J., Ding, W. & Lin, C. ChatGPT: tackle the growing carbon footprint of generative AI. *Nature* **615**, 586-586 (2023).

47 Patterson, D. *et al.* Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350* (2021).

48 Li, P., Yang, J., Islam, M. A. & Ren, S. Making AI Less" Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. *arXiv preprint arXiv:2304.03271* (2023).

49 Microsoft. Environmental Sustainability Report. (2022).

50 Nordgren, A. Artificial intelligence and climate change: ethical issues. *Journal of Information, Communication and Ethics in Society* **21**, 1-15 (2022).

51 Nabavi, E. *ChatGPT's Impact on Academia: A Canary in the Coalmine of a Rapidly Changing Landscape*, <https://t.ly/85D77> (2023).

52 SyntheticUsers. *Synthetic Users: Merging Qualitative and Quantitative Research, in seconds*, <https://www.syntheticusers.com/post/synthetic-users-merging-qualitative-and-quantitative-research> (2023).

53 Raman, S. Science, Uncertainty, and the Normative Question of Epistemic Governance in Policymaking. *Knowledge, Technology and Law*, 17-32 (2014).

54 Pearce, W., Mahony, M. & Raman, S. Science advice for global challenges: Learning from trade-offs in the IPCC. *Environmental Science & Policy* **80**, 125-131 (2018). https://doi.org:https://doi.org/10.1016/j.envsci.2017.11.017

55 Jasanoff, S. Serviceable truths: Science for action in law and policy. *Tex. L. Rev.* **93**, 1723 (2014).

56 Gluckman, P. D., Bardsley, A. & Kaiser, M. Brokerage at the science–policy interface: from conceptual framework to practical guidance. *Humanities and Social Sciences Communications* **8**, 84 (2021). https://doi.org:10.1057/s41599-021-00756-3

57 Zhou, J., Zhang, Y., Luo, Q., Parker, A. G. & De Choudhury, M. in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1-20.

58 Bagdasaryan, E. & Shmatikov, V. in *2022 IEEE Symposium on Security and Privacy (SP).* 769-786 (IEEE).

59 Lubchenco, J. Entering the century of the environment: a new social contract for science. *Science* **279**, 491-497 (1998).

60 Crow, J. M. Achieving UN climate goals needs purposeful, persistent action from science. *Nature* (2023).

61 Nabavi, E. & Browne, C. Leverage zones in Responsible AI: towards a systems thinking conceptualization. *Humanities and Social Sciences Communications* **10**, 1-9 (2023).