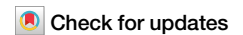


<https://doi.org/10.1038/s43247-024-01392-w>

Generative AI tools can enhance climate literacy but must be checked for biases and inaccuracies



Carmen Atkins ^{1,2}, Gina Girgente ³, Manoochehr Shirzaei ^{1,2,4} & Junghwan Kim ⁵ ✉

In the face of climate change, climate literacy is becoming increasingly important. With wide access to generative AI tools, such as OpenAI's ChatGPT, we explore the potential of AI platforms for ordinary citizens asking climate literacy questions. Here, we focus on a global scale and collect responses from ChatGPT (GPT-3.5 and GPT-4) on climate change-related hazard prompts over multiple iterations by utilizing the OpenAI's API and comparing the results with credible hazard risk indices. We find a general sense of agreement in comparisons and consistency in ChatGPT over the iterations. GPT-4 displayed fewer errors than GPT-3.5. Generative AI tools may be used in climate literacy, a timely topic of importance, but must be scrutinized for potential biases and inaccuracies moving forward and considered in a social context. Future work should identify and disseminate best practices for optimal use across various generative AI tools.

The reality of climate change is quickly becoming apparent^{1,2} as an increasing number of people are experiencing the various impacts of climate change through hazards, such as droughts and floods, which are generally expected to intensify in both magnitude and frequency³⁻⁵. The impacts of climate change-related hazards are wide-ranging and yet to be fully recognized^{6,7}. Therefore, climate change represents an increasingly relevant topic in many academic fields and global society as a whole⁸. Adaptation to and mitigation of climate change and its impacts requires that policymakers, researchers, and engaged citizen stakeholders develop and maintain climate literacy in order to plan ahead and implement adaptations efficiently.

Climate literacy refers to the capacity to synthesize information regarding the climate within varying contexts⁹. Not only researchers and policymakers but also ordinary citizens can benefit from climate literacy. There are multiple reasons for this. First, with adequate climate literacy, individuals can discern the meaning and credibility (or lack of credibility) behind news articles. Moreover, individuals can sufficiently respond to both the economic and environmental ramifications of climate change and apply knowledge of climate change to their careers, as such knowledge impacts a vast array of fields¹⁰. In countries like the United States, citizens vote and pay taxes—many of which are relevant to sustainable policies and disaster response plans. If lacking climate literacy, individuals and the organizations and governments they make up may underestimate the urgency of adaptation measures to climate change, waiting to respond until it is too late to avoid the most damaging effects¹¹.

The importance of climate literacy is underscored by younger generations, such as Generation Z, who comprise the future stakeholders that will formulate policies and actions, which will either exacerbate or mitigate the negative impacts of climate change around the globe¹². Kuthe et al.¹² identify teenagers as a target demographic of top priority since they will be the ones to take on the hazards of climate change, placing much of the future's environmental conditions in their hands. Moser¹³ emphasizes this by identifying gaps in the public's understanding of climate change and related issues. Clearly, climate literacy should be prioritized, especially focusing on younger individuals, for they represent the highest potential for implementing actions that will favorably impact the future¹⁴.

Considering the importance of enhancing climate literacy, the recent advances in generative artificial intelligence (AI), such as OpenAI's ChatGPT and Google's Bard, may hold meaningful implications for climate literacy. Such generative AI tools are expected to provide a more effective means to obtain new knowledge and information than conventional methods based on web search engines¹⁵⁻¹⁷, although it should be noted that these tools are not unanimously available. For example, Google's Bard is not legal in China, nor available in Canada¹⁸. Additionally, younger generations comprise those who will be most affected by climate change, as well as those who will carry out any adaptation measures—effectively placing many of the practical outcomes of climate change in the hands of our youth and future generations^{19,20}. This, combined with the growth of nontraditional learning platforms and tools²¹, has stimulated ongoing discussions among educators

¹Virginia Tech, Department of Geosciences, Blacksburg, VA, USA. ²Virginia Tech, Virginia Tech National Security Institute, Blacksburg, USA. ³Independent Researcher, Richmond, VA, USA. ⁴United Nations University, Institute for Water, Environment and Health, Hamilton, Ontario, Canada. ⁵Virginia Tech, Department of Geography, Blacksburg, VA, USA. ✉e-mail: junghwankim@vt.edu

about the potential role of generative AI platforms in learning environments^{22–24}. In other words, generative AI tools are expected to become essential tools for students and younger generations to improve their climate literacy.

With this growing need to improve the climate literacy of younger generations and the increasingly common use of generative AI platforms, we argue that researchers should examine the potential capabilities and weaknesses (e.g., inaccuracies and biases) of these tools^{15–17}, particularly in the context of climate change topics^{25,26}. Without acknowledging the weaknesses of specific AI tools, students may falsely believe such tools function without error, resulting in false information provided to students who integrate it into their understanding of climate change. This may eventually lead to severe educational problems, such as hallucination effects^{27,28}. On the other hand, if generative AI platforms are shown to be sufficiently reliable, they may be used as an accessible means to enhance climate literacy. Our study takes an exploratory approach to this timely issue that, to the best of our knowledge, has yet to be addressed by previous studies.

We select OpenAI's ChatGPT as our case study (using both GPT-3.5 and GPT-4). While many generative AI tools exist, with many more expected to be released in the near future, we choose to focus on ChatGPT for this case study for the following reasons: First, ChatGPT has experienced the most drastic acceleration of usage since its release²⁹ and second, individuals aged 18–34 currently comprise over 60% of ChatGPT use³⁰. Third, ChatGPT represents a prominent tool and shows an early adoption of usage in developing nations, for which many uses of AI can be identified³¹. While each specific AI tool should be examined with the same questions in mind, we focus here on ChatGPT as an initial exploration into the issue of climate change literacy and generative AI.

Overall, this study aims to examine the accuracy of ChatGPT's responses to climate change-related hazards across the globe by comparing responses to credible hazard risk indices, which are based on data used in the International Panel on Climate Change (IPCC) Annual Review 6 (AR6), Working Group II, Chapter 8³². We find overall agreement between ChatGPT responses and the hazard risk indices for floods and cyclones, but lower agreement regarding droughts, as well as improved consistency and reduced errors for GPT-4 responses (in comparison to GPT-3.5). This study offers an empirical attempt to systematically investigate the general capabilities and weaknesses of ChatGPT (as of December 2023) regarding country-level vulnerabilities to climate change-related hazards.

Results

Number of climate change-related topics by GPT-4

The topic counts per country from the first iteration of GPT-4 responses are displayed in Fig. 1(a), demonstrating the spatial variation across continents. On average, 9,089 topics are identified with a standard deviation of 1.129. The minimum value is 6, while the maximum value is 12. We further analyze the consistency in ChatGPT's responses. Figure 1(b) shows the standard deviation of the number of topics per country over the 10 iterations. While the topic count variation remains fairly low for GPT-4 across each continent, many countries in Africa and some countries in the Middle East seem to have the least consistency, suggesting that ChatGPT's responses are relatively less consistent in these regions compared to other regions.

Accuracy of GPT-4 results compared with the IPCC report

Overall, the first iteration of results created by GPT-4 proved fairly accurate compared to the validation data. Recall that three climate change-related hazard issues were selected for accuracy analysis—floods, droughts, and cyclones—because they create extensive, yet different, damage, thus needing to be monitored as climate change continues. Table 1 shows confusion matrices for each issue across one iteration.

Cyclone themes were the most accurate, with an accuracy score of 0.806. This means GPT-4 accurately identified cyclones as a climate change-related hazard 80.6% of the time. 20 false negatives and 17 false positives

were produced for this theme. Flooding was accurately mentioned 76.4% of the time. False negatives and false positives for flooding were relatively of the same frequency, with counts of 20 and 25, respectively. There was no substantial difference between false positives and false negatives for floods and cyclones. However, while still having a reasonable accuracy score, droughts were the topic that GPT-4 struggled with most. Droughts were accurately identified 69.1% of the time, and there were 17 more false negatives than false positives.

Additionally, we examine how accuracy scores change for each hazard across 10 iterations, as seen in Fig. 1(c). Specifically, flood accuracy has an 8-percentage point difference with a range of 0.743–0.822, drought accuracy has a 5-percentage point difference with an accuracy range of 0.639–0.651, and cyclone accuracy has a 5-percentage point difference with an accuracy range from 0.770–0.822 across the 10 iterations. Overall, we conclude that accuracy scores for floods, droughts, and cyclones are consistent across all GPT-4 iterations.

Comparison between GPT3.5 and GPT-4 models

Overall, GPT-3.5 (default model, Fig. 2[a]) seems less reliable than GPT-4, which is the most advanced model. For instance, GPT-3.5 showed limited abilities to correctly produce responses in accordance with our prompt directions. As seen in Fig. 2(b), out of 1,910 prompt requests (i.e., 191 countries × 10 iterations), 38 outputs from GPT-3.5 were in the incorrect format, which did not allow us to further process the outputs. Note that GPT-4 did not have issues providing the correct format for each of the 1,910 cases, suggesting that it is a more capable and reliable tool than GPT-3.5, at least in this regard.

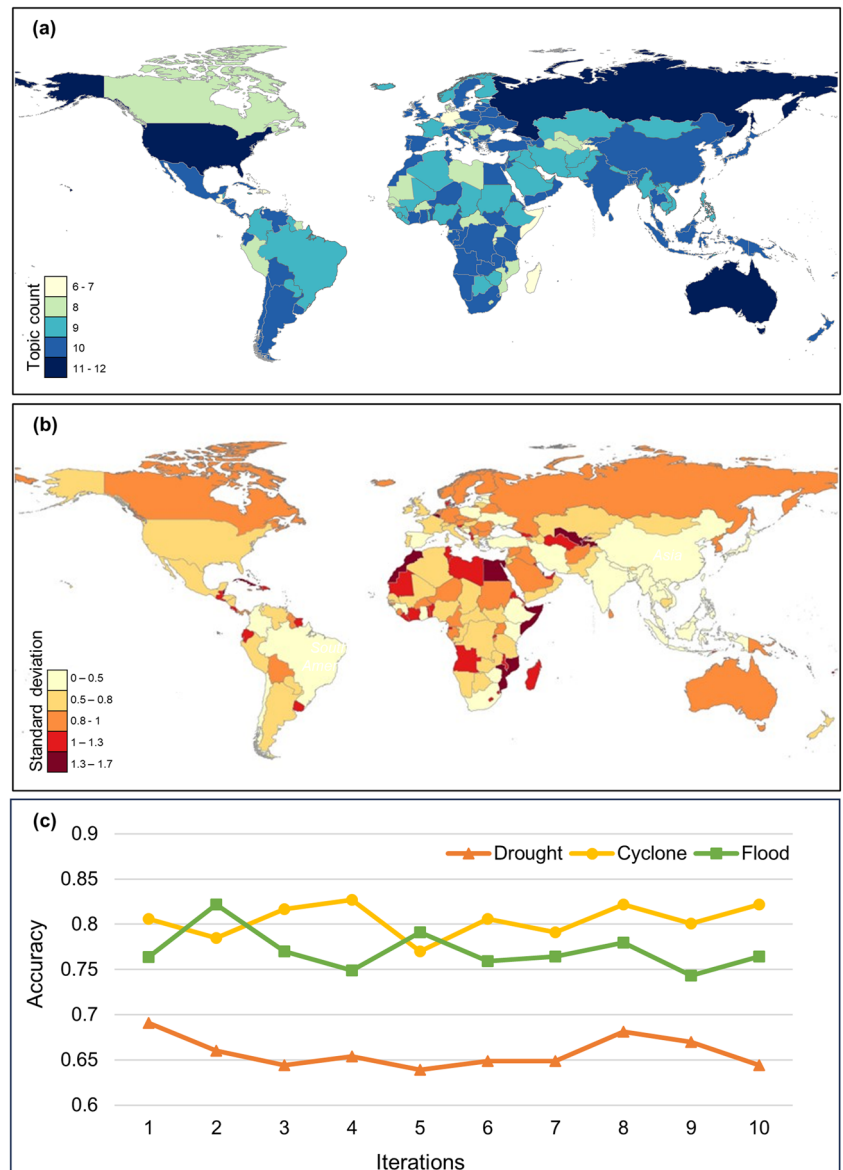
In light of this, assuming that GPT-4's responses are reliable and accurate sources, Fig. 2b reports many errors, particularly for countries residing in Africa, with the general Europe-Asia boundary area being the second highest in errors. Perhaps the largest difference regarding topic counts between the two models can be identified in South America and Ireland (Fig. 2a). However, in general, all continents appear to have a similar distribution. Regarding the descriptive statistics of results obtained from GPT-3.5, the average number of topics identified by GPT-3.5's responses is 9.426 with a standard deviation of 1.249. The minimum value of topic counts is 6, while the maximum value of topic counts is 15. The paired sample *t*-test results indicate that there is a statistically significant difference ($p < 0.01$) in the number of identified topics by GPT-3.5 and GPT-4.

Discussion

By focusing on ChatGPT as a case study, our exploratory study achieves one of the first steps toward informing users of generative AI tools' potential strengths and weaknesses relevant to climate change literacy. By comparing three major hazards (floods, droughts, and cyclones) reported for each country by ChatGPT and comparing each to the validation data, we identified more accuracies than inaccuracies in ChatGPT's responses—but not enough to conclude that the tool, when used in this way, is truly reliable. For example, ChatGPT tends to underestimate vulnerability to droughts, as ChatGPT reports droughts as a primary risk for considerably fewer countries than the trusted validation data do. This presents a false negative type error, which may potentially mislead ChatGPT's users, who are currently formulating a sense of security and severity. For floods and cyclones, however, the opposite is true: most inaccuracies stem from false positives. Depending on the hazard, these trends in false positives/negatives present important biases and limitations that users should be aware of.

Despite the inaccuracies both types (false positive and false negative) clearly present, a considerable level of agreement is found between the ChatGPT responses and validation data for cyclones and floods. This is confirmed by the high accuracy scores across the 10 iterations of the GPT-4 model. However, the results also report a relatively lower level of agreement for droughts, as evidenced by lower accuracy scores than the other two hazard cases. Overall, our results suggest that, although the false positive bias should be kept in mind, ChatGPT may be used—with caution—as a starting point for users looking to gain climate literacy regarding some hazards, like

Fig. 1 | Topic count and accuracy results for GPT-4. a Spatial variation of topic counts across the first GPT-4 iteration, where topic count increases from light green to dark blue. **b** Map of the standard deviation of topic counts for GPT-4 across all ten iterations where standard deviation increases from light yellow to dark red. **c** Accuracies for droughts (orange triangles and line), cyclones (yellow circles and line), and floods (green squares and line) for GPT-4. Maps and graphs were created by authors.



floods and cyclones. However, considering droughts, more caution should be employed, as false negatives are arguably more dangerous in this context and overall accuracy is lower.

One should naturally ask what the origins of these inaccuracies might be. While identifying true causes is beyond the scope of this exploratory study, we suggest a few possible factors that may influence the performance of ChatGPT in this context. First, we must consider that this study was conducted entirely in English. As OpenAI has acknowledged, a bias toward English and perspectives aligning with Western cultures exists in the AI³³. This bias may be relevant both to the responses generated by ChatGPT, which cater to Western, English-speaking users, as well as the AI’s processing of prompts—i.e., it may comprehend prompts from native English-speakers best. This situation is especially important to consider for regions within the Global South, where climate literacy is an important, yet poorly understood issue³⁴. Perceptions of climate change risk vary widely across different cultures³⁵, making even small semantic changes in ChatGPT responses potentially impactful. This language-related bias—in both ChatGPT functioning and user experience—introduces an additional variable to consider, the effects of which are not yet fully understood and may account for general variation in results, if this study were repeated in a non-English language. Additionally, regarding the lower accuracy for

droughts (as compared to floods and cyclones), we must consider how such hazards are defined. The IPCC itself has acknowledged that drought is a relative term³⁶, depending on many factors and contexts. Definitions in various sources other than the IPCC and related data sources may, therefore, vary more than other hazards like cyclones, which are more prominent and transparent in definitions (i.e., there is no debate over ongoing cyclones). This could partly explain the decreased accuracy in our validation of droughts, as opposed to floods and cyclones. Overall, this issue related to the definitions of hazards might contribute to the uncertainties of our analytical results, which future studies can examine through sensitivity analyses.

While not completely accurate compared to the validation data, GPT-4 offers a suggested pattern of consistency and reliability in its output regarding topic counts across 10 iterations. However, GPT-3.5 demonstrates unreliability as it produces errors when creating its responses, which we never encountered with GPT-4. Therefore, if possible, our results recommend that users employ GPT-4 rather than GPT-3.5. While it is unsurprising that GPT-4—the more advanced and costly version—performs better than the default version (GPT-3.5), this suggests potential ethical issues regarding tools available to users of different socioeconomic positions^{37–39}. These potential ethical concerns can be especially relevant

Table 1 | Confusion matrices for floods, droughts, and cyclones

Validation (IPCC)			
GPT-4	Topic: Floods (Accuracy: 0.764)	Vulnerable to floods	NOT vulnerable to floods
	Vulnerable to floods	3 (TP)	25 (FP)
	NOT vulnerable to floods	20 (FN)	143 (TN)
GPT-4	Topic: Droughts (Accuracy: 0.691)	Vulnerable to droughts	NOT vulnerable to droughts
	Vulnerable to droughts	32 (TP)	21 (FP)
	NOT vulnerable to droughts	38 (FN)	100 (TN)
GPT-4	Topic: Cyclones (Accuracy: 0.806)	Vulnerable to cyclones	NOT vulnerable to cyclones
	Vulnerable to cyclones	3 (TP)	17 (FP)
	NOT vulnerable to cyclones	20 (FN)	151 (TN)

This table reports the results of GPT-4's first iteration. True positives (TP) occur when both GPT-4 and validation data identify the theme as a vulnerability for the given country, whereas true negatives (TN) occur when both sources do not claim a theme as an issue for the selected country. False negatives (FN) are established when ChatGPT does not state a theme to be an issue while the validation data do. Conversely, false positives (FP) are shown when GPT-4 inappropriately claims a theme to be an issue for a country while the validation data do not.

considering that those in developing economies have been some of the fastest populations to adopt applications of ChatGPT³¹.

We believe that a comprehensive examination of the capabilities of generative AI tools, such as ChatGPT and Bard, will likely grow in value, considering their quickly increasing role in climate literacy^{35,26} and their potential—yet debated—beneficial applications in the general education sector^{22,27,40} within countries where it is available. While providing insight into generative AI's ability to summarize climate change-related hazards on a global, country-level scale, our study contains limitations that should not be overlooked. By utilizing the default parameters and API service that was initiated at each iteration, we provide data that, to the best of our knowledge, is minimally influenced by the user's prompt history^{16,41}. However, because of the black-box nature of AI models⁴², it must be noted that individual users may experience different outputs. Further, we recommend that future studies consult OpenAI's documentation for relevant updates to either GPT version (since December 2023), as OpenAI regularly updates each model. Another variable to consider is that of user demand—might the performance of either version, particularly GPT-3.5, degrade with increased user demand at a given time? Next, we must also consider the limitations of the BERT NLP processing model with which we consolidated the ChatGPT responses into 50 themes. While the NLP model allows us to automate the consolidation process and reduce human error, and we employed the Davies-Bouldin Index, Silhouette, and Within-Cluster Sum of Squares scores (Supplementary Fig. 1), BERT is not perfect, and minor errors in clustering are possible, such as a group of temperature change topics including the more general topic of 'arctic change.' However, because BERT takes context into account, such an example may have related to temperature change in the original text. Regardless, this reminds us that BERT functions as a 'black box' model, which leaves us with unknowns that, for the time being, we simply accept. Keeping this in mind, we state that, within reasonable feasibility, the BERT model still offers improvements to this study's approach in accuracy (eliminating human error) and efficiency (completing the same job manually would be nearly impossible, requiring contextual analysis of thousands of topics). Therefore, considering the high sample rate, we conclude that sparse random errors are acceptable for the scope of our study, especially in comparison to a manual approach. Thus, future studies are recommended to mitigate these uncertainties and limitations of the NLP model to provide a more robust theme consolidation

result. Finally, the likely bias relating to the English language should be considered in additional cases⁴³.

Further work should continue to comprehensively investigate the performance of the many additional emerging generative AI tools, such as Google's Bard and ChatClimate (www.chatclimate.ai/)—a customized large language model developed by researchers²⁶ for climate literacy-related use. Future studies are also recommended to quantify the limitations of these tools as precisely and comprehensively as possible. Potential geographic biases resulting from training datasets should also be examined more quantitatively^{16,44,45}. One potential means to conduct further investigation into this issue would be to conduct a Delphi study⁴⁶, which could offer insights before a wealth of established literature is available. Finally, developing educational recommendations for potential users of these AI tools is essential. More studies are being published, which indicate that prompt engineering and parameter-setting for GPT-4 are key for utilizing the tool effectively¹⁶. In light of this, we recommend further studies to examine the factors discussed here and develop best-practice guidelines. While most studies now focus on GPT-4 and its many additional capabilities, it is important to inform users of biases present in GPT-3.5, as many users, especially non-academic, will still use only the default version. This study puts forth an overview of country-level vulnerabilities to climate change-related hazards as told by both versions of ChatGPT as of December 2023.

In conclusion, climate change adaptation strategies will be dependent on the upcoming generations and their climate literacy—people's understanding of climate change and willingness to be involved in mitigation and adaptation. This is a crucial point in the future of our planet, as projections show that waiting any longer to reduce climate emissions may result in a point of irreversible consequences⁴⁷. Moreover, considering the growing importance of generative AI tools and their uptake by individuals worldwide, future studies on the combined topic of generative AI tools and climate literacy should commence with the ultimate goal of disseminating findings to enable informed, discerning use of ChatGPT and other increasingly popular generative AI platforms toward the pressing issue of climate change.

Methods

ChatGPT prompting

We performed analyses for both GPT-3.5, the default version, and GPT-4, the advanced version, to monitor any potential response differences between the two. Figure 3 illustrates an overview of the research methods. We formulated a prompt template (Supplementary Note 1) to inquire about a country's vulnerability to climate change-related hazards. For example, we used the following prompt (similar to that in Kim et al.¹⁶) as our input to investigate Australia: "List the climate change-related hazards that Australia is most vulnerable to. Provide a numbered list of the climate change-related hazards with descriptions. Make sure to put a colon between the numbered list and the description. The listed climate change-related hazards should not be duplicated with each other." This prompt was submitted for the 191 IPCC member countries. To examine to what extent ChatGPT's responses are consistent in terms of different experiments, the prompt was repeated ten times for each country and each ChatGPT version. In total, 4,018 topics were created by both ChatGPT versions.

To process these data effectively, we used OpenAI's ChatGPT application programming interface (API)⁴³; see Supplementary Note 2. Following an approach by Kim et al.¹⁶, we instructed the system to act as a helpful assistant and then began using our prompt. Notice that our prompt template instructs ChatGPT to report the hazards in a list form and place a colon after each topic name is introduced, thus allowing the API to extract all topics accurately and automatically from each response¹⁶. Regarding parameters that might affect ChatGPT's outputs, we used all default settings, including that for temperature (randomness of responses) and max_tokens (response length). Responses per country took approximately 30 seconds to retrieve, on average.

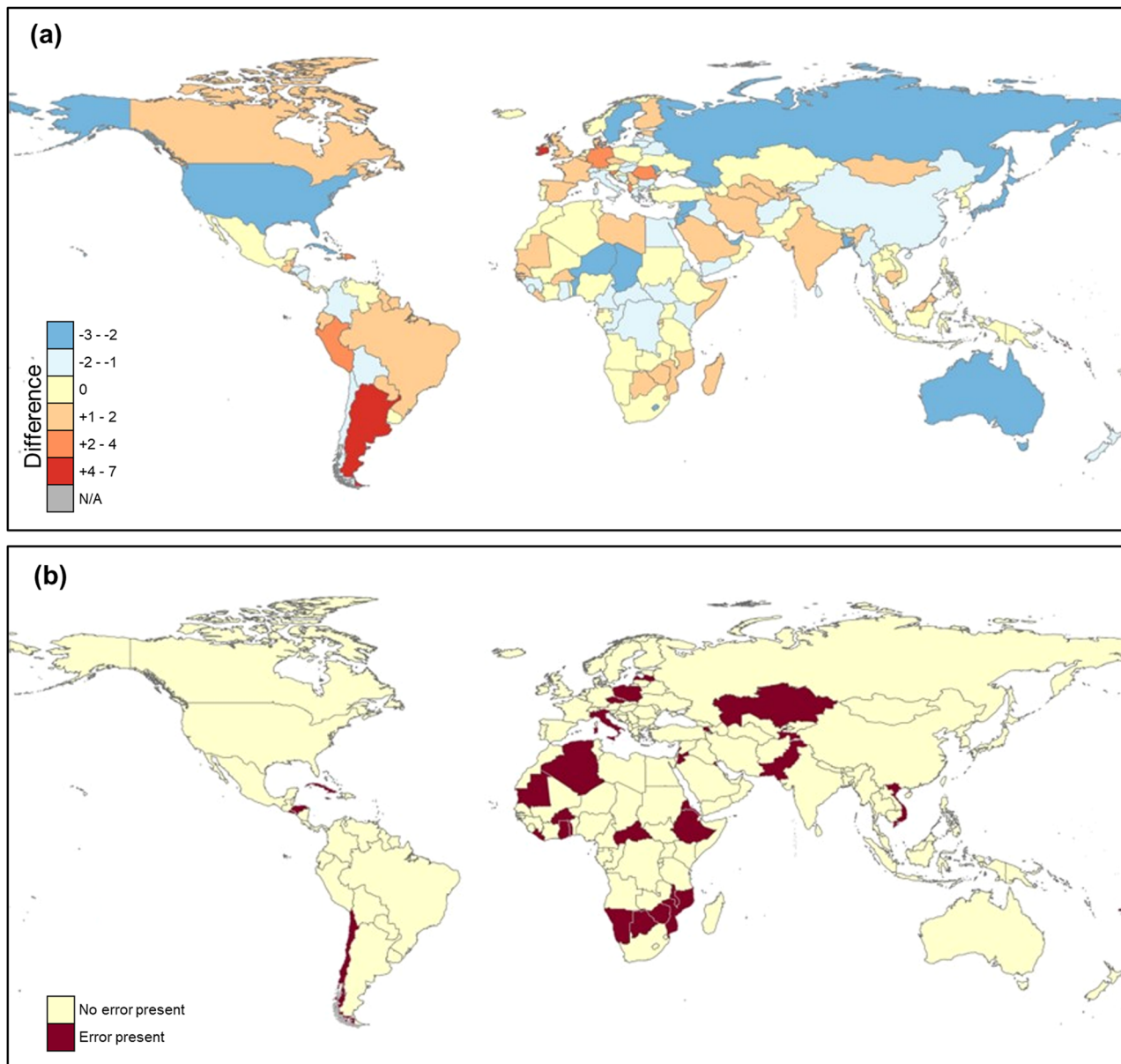


Fig. 2 | Topic count differences between models and general error maps. **a** Map showing the difference between topic count numbers (GPT-3.5 topic count – GPT-4 topic count). **b** Map showing the countries where output errors (dark red) occurred

in GPT-3.5; light yellow indicates countries for which no error was observed in output responses. Maps were created by authors.

Topic consolidation

To consolidate the topics (4018 once duplicates were removed) into similar topic clusters that are meaningful to be used for analysis, we employed the Bidirectional Encoder Representations from Transformers (BERT), a natural language processing (NLP) model⁴⁸; see Supplementary Note 3. BERT is an open-source model that incorporates the context behind a word by comparing it to all other words in a sentence^{48–50}. This capability allows it to efficiently identify recurring topics mentioned in responses from ChatGPT. K-means ++ clustering^{51–54} was used to reduce the 4,018 unique topics into 50 topic clusters, which we refer to as themes. We identified 50 clusters as the optimal number of clusters by referring to the Davies-Bouldin Index, the within-cluster sum of squares (WCSS), and silhouette scores¹⁶; see Supplementary Fig. 1. From these results, we obtain basic descriptive statistics for GPT-3.5 and GPT-4. We also perform a paired sample *t*-test on topic counts to test if the difference between the identified topics of the two GPT versions is significant.

Comparison with validation data

To explore the accuracy and consistency of ChatGPT responses, we performed data validation by applying the Index for Risk Management (INFORM) Global Risk Index (GRI)^{55,56} to the ChatGPT results. The INFORM data set is based on the data within Chapter 8 of the most recent IPCC AR6 Working Group II³¹ and provides widely accepted, comprehensive measures of risk due to climate-related factors (including hazards), which are offered at the country-level scale. We use the 2019 version, which would, therefore, have been available to be used as training data for ChatGPT. With such reasons in mind, we chose these data for our validation process and hereafter refer to them as the validation data. We used this dataset to validate the ChatGPT responses for floods, droughts, and cyclones —three major climate change-related hazards that were included in both the validation data and ChatGPT response themes. The validation data consists of rankings from “very low” to “very high.” In order to compare these indices with our binary classification of the ChatGPT data, we translated the indices

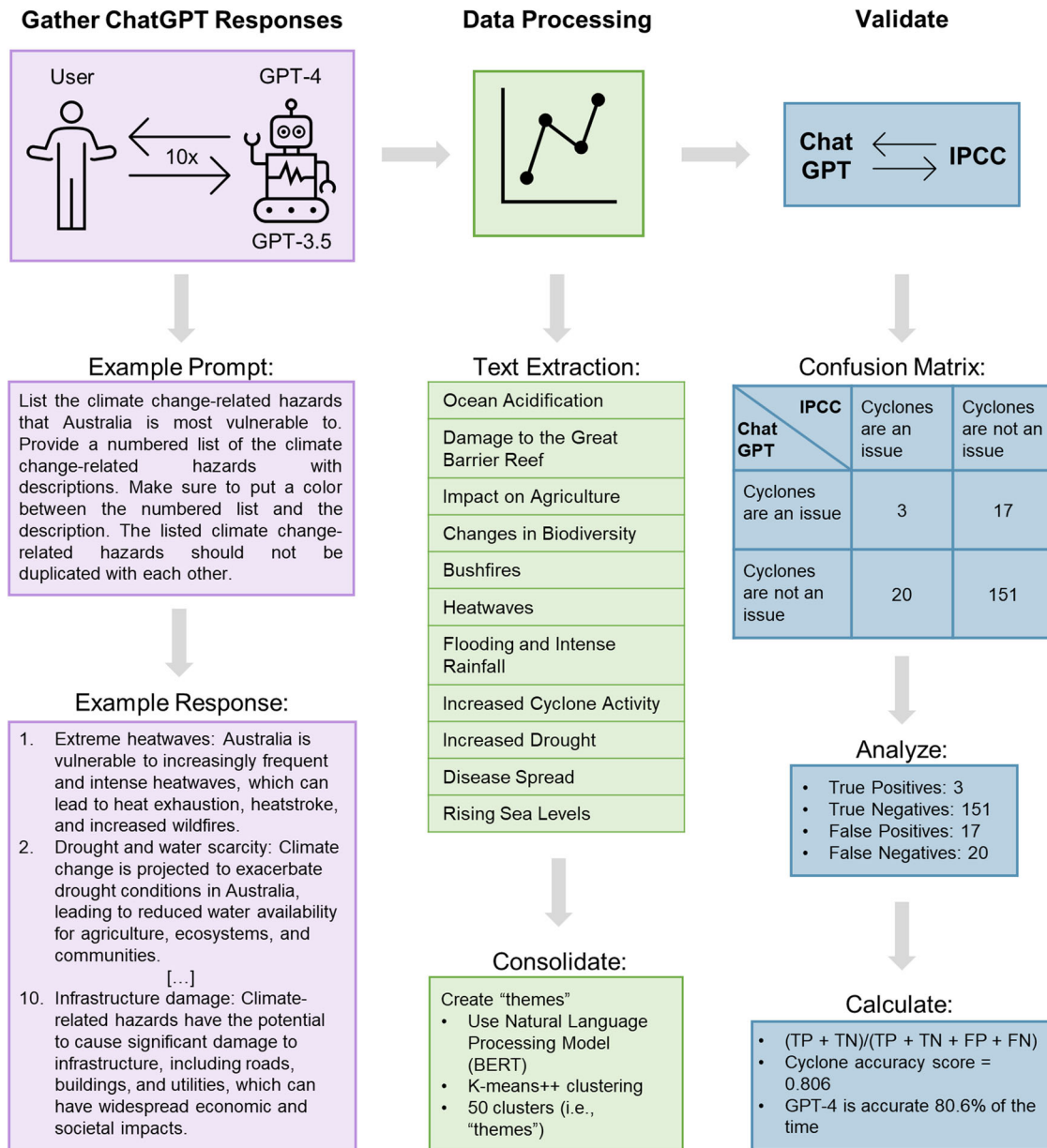


Fig. 3 | Project workflow. Overview of the research methods. The top row and titles over each cell indicate the main components of our methods, while the respective columns provide more detailed steps and examples; the colors are only for distinguishing columns. The workflow diagram was created by authors.

for “medium,” “high,” and “very high” as value 1 (i.e., presented), with anything below medium risk as value 0 (i.e., not presented).

We created confusion matrices between the validation data and ChatGPT responses for each validation hazard. The confusion matrices specify which hazard vulnerabilities the two sources (i.e., ChatGPT responses and validation data) agreed on (true positive [TP] and true negative [TN]) or disagreed on (false positive [FP] and false negative [FN]).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

By combining the accuracy scores, and basic descriptive statistics for each iteration of responses from GPT-3.5 and GPT-4, we observe if there is a theme of general consistency between iterations and quantify the level of (dis)agreement between ChatGPT and the IPCC.

Data availability

The data necessary to replicate this study, as well as consolidated topics (results) can be found at (https://figshare.com/articles/dataset/BERT_Consolidated_Topics/25403407), with data used for validation found at (<https://sedac.ciesin.columbia.edu/data/set/ipcc-inform-gri-2019-v0-3-7>). Please note: Data gathered from ChatGPT (chat.openai.com) will inherently vary over time, as publicly available models change. The ChatGPT responses in this study were acquired using models GPT-3.5 and GPT-4, accessed in October 2023.

Code availability

The code necessary to replicate this study can be found at (https://figshare.com/articles/dataset/BERT_Consolidated_Topics/25403407).

Received: 9 August 2023; Accepted: 15 April 2024; Published online: 30 April 2024

References

- IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Masson-Delmotte, V. et al.) 2391 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021). <https://doi.org/10.1017/9781009157896>.
- Steffen, W. et al. Trajectories of the earth system in the anthropocene. *Proc. Natl. Acad. Sci.* **115**, 8252–8259 (2018).
- Milly, P. et al. Increasing risk of great floods in a changing climate. *Nature* **415**, 514–517 (2002).
- Naumann, G. et al. Global changes in drought conditions under different levels of warming. *Geophys. Res. Lett.* **45**, 3285–3296 (2018).
- Pokhrel, Y. et al. Global terrestrial water storage and drought severity under climate change. *Nat. Clim. Chan.* **11**, 226–233 (2021).
- Carleton, T. A. & Hsiang, S. M. Social and economic impacts of climate. *Science* **353**, aad9837 (2016).
- Steffen, W. et al. Planetary boundaries: Guiding human development on a changing planet. *Science* **347**, 1259855 (2015).
- Nalau, J. & Verrall, B. Mapping the evolution and current trends in Climate change adaptation science. *Clim. Risk Manage.* **32**, 100290 (2021).
- UNESCO. The UNESCO Climate Change Initiative (2010) Climate change education for sustainable development, Paris Futerra (1996) Sell the Sizzle. *New Clim. Message, UNESCO ED.2010/WS/41* (2010).
- U.S. Global Change Research Program (Ed.). *Climate literacy: The essential principles of climate sciences: A guide for individuals and Communities. U.S. Global Change Research Program.* (2009).
- Eitzinger, A., Binder, C. R. & Meyer, M. A. Risk perception and decision-making: do farmers consider risks from climate change? *Climatic Change* **151**, 507–524 (2018).
- Kuthe, A. et al. How many young generations are there? – a typology of teenagers' climate change awareness in Germany and Austria. *J. Environ. Educ.* **50**, 172–182 (2019).
- Moser, S. C. Reflections on climate change communication research and practice in the second decade of the 21st Century: What more is there to say? *WIREs Clim. Change* **7**, 345–369 (2016).
- Syropoulos, S. & Markowitz, E. Our responsibility to future generations: the case for intergenerational approaches to the study of Climate Change. *J. Environ. Psychol.* **87**, 102006 (2023).
- Kim, J. & Lee, J. How does ChatGPT Introduce Transport Problems and Solutions in North America? *Findings*, 1–6 <https://doi.org/10.32866/001c.72634> (2023).
- Kim, J., Lee, J., Jang, K. M. & Lourentzou, I. Exploring the limitations in how CHATGPT introduces environmental justice issues in the United States: A case study of 3,108 counties. *Tele. Inform.* **86**, 102085 (2024).
- Voß, S. Bus Bunching and Bus Bridging: What Can We Learn from Generative AI Tools like ChatGPT? *Sustainability* **15**, 9625 (2023).
- Google. Where you can use Bard. *Bard Help*. <https://support.google.com/bard/answer/13575153?hl=en>. Accessed January 2024. (2024).
- Sanson, A. V., Van Hoorn, J. & Burke, S. E. Responding to the impacts of the climate crisis on children and Youth. *Child Dev. Perspect.* **13**, 201–207 (2019).
- Skeiryte, A., Krikstolaitis, R. & Liobikienė, G. The differences of climate change perception, responsibility and climate-friendly behavior among generations and the main determinants of Youth's climate-friendly actions in the EU. *J. Environ. Manage.* **323**, 116277 (2022).
- Clayton, K., Blumberg, F. & Auld, D. P. The relationship between motivation, learning strategies and choice of environment whether traditional or including an online component. *Br. J. Educ. Technol* **41**, 349–364 (2010).
- Kung, T. H. et al. Performance of chatgpt on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit. Health* **2**, e0000198 (2023).
- Stokel-Walker, C. Ai Bot chatgpt writes smart essays — should professors worry? *Nature*. <https://doi.org/10.1038/d41586-022-04397-7> (2022).
- Vogels, E. A majority of Americans have heard of ChatGPT, but few have tried it themselves. *Pew. Res.* <https://www.pewresearch.org/short-reads/2023/05/24/a-majority-of-americans-have-heard-of-chatgpt-but-few-have-tried-it-themselves/> (2023).
- Larosa, F. et al. Halting generative AI advancements may slow down progress in climate research. *Nat. Clim. Change* **13**, 497–499 (2023).
- Vaghefi, S. A. et al. ChatClimate: Grounding conversational AI in climate science. *Commun. Earth Environ.* **4**, 480 (2023).
- Day, T. A preliminary investigation of fake peer-reviewed citations and references generated by Chatgpt. *Prof. Geograph.* **1–4**, <https://doi.org/10.1080/00330124.2023.2190373> (2023).
- Shen, Y. et al. ChatGPT and other large language models are double-edged swords. *Radiology* **307**, e230163 (2023).
- UNESCO. *Guidance for generative AI in education and research*, (2023). 978-92-3-100612-8.
- Turner, A. *30 CHATGPT User & Market Size Statistics*. BankMyCell. <https://www.bankmycell.com/blog/chatgpt-number-of-users#4>. Accessed December 2023. (2023).
- Kshetri, N. Chatgpt in developing economies. *IT Professional* **25**, 16–19 (2023).
- Birkmann, J. E. et al. 2022: poverty, livelihoods and sustainable development. In *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Pörtner, H.-O. et al.) 1171–1274 (Cambridge University Press, 2022).
- OpenAI. *Is CHATGPT biased?* <https://help.openai.com/en/articles/8313359-is-chatgpt-biased> Accessed December 2023. (2023).
- Simpson, N. P. et al. Climate change literacy in Africa. *Nat. Clim Change* **11**, 937–944 (2021).
- Lee, T. et al. Predictors of public climate change awareness and risk perception around the world. *Nat. Clim Change* **5**, 1014–1020 (2015).
- IPCC. Glossary of terms. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)* (eds Field, C. B. et al.) 555–564 (Cambridge University Press, 2012).
- Kim, D., Zhu, Q. & Eldardiry, H. Exploring approaches to artificial intelligence governance: from ethics to policy. In *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)* 1–5 (IEEE, 2023).
- Mbakwe, A. B., Lourentzou, I., Celi, L. A., Mechanic, O. J. & Dagan, A. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLOS Digital Health* **2**, e0000205 (2023).
- Ruane, E., Birhane, A. & Ventresque, A. Conversational AI: Social and Ethical Considerations. In *AICS* **2563**, 104–115 (2019).
- Mbakwe, A. B., Lourentzou, I., Celi, L. A. & Wu, J. T. Fairness metrics for health AI: we have a long way to go. *Ebiomedicine* **90**, 104525 (2023).
- Kosinski, M. Theory of mind may have spontaneously emerged in large language models. Preprint at <https://doi.org/10.48550/arXiv.2302.02083> (2023).
- Hu, Y. et al. GeoAI at ACM SIGSPATIAL: progress, challenges, and future directions. *Sigspatial Spec.* **11**, 5–15 (2019).
- OpenAI. *OpenAI API* <https://openai.com/blog/openai-api> (2023). Accessed December 2023.
- Graham, M., Hogan, B., Straumann, R. K. & Medhat, A. Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Ann. Assoc. Am. Geograph.* **104**, 746–764 (2014).

45. Jang, K. M. et al. Understanding place identity with generative AI (short paper). In *12th International Conference on Geographic Information Science (GIScience 2023)* 41:1–41:6 (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023).
 46. Niederberger, M. & Spranger, J. Delphi technique in Health Sciences: A Map. *Front. Public Health* **8**, 457 (2020).
 47. McKay, A. D. et al. Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science* **377**, eabn7950 (2022).
 48. Devlin, J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *North American Chapter of the Association for Computational Linguistics* **2019**, 4171–4186 (2019).
 49. Reimers, Nils, and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv* <https://doi.org/10.48550/arXiv.1908.10084> (2019). 1908.10084.
 50. Wolf, T. et al. Transformers: state-of-the-art natural language processing. In *Proc. 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* 38–45 (Association for Computational Linguistics, 2020).
 51. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
 52. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Engineer.* **9**, <https://doi.org/10.1109/MCSE.2007.55> (2007).
 53. Reback, J. et al. pandas-dev/pandas: Pandas 1.0. 5. *Zenodo* <https://doi.org/10.48550/arXiv.1908.10084> (2020).
 54. Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **17**, 261–272 (2020).
 55. Joint Research Centre - JRC - European Commission. INFORM Global Risk Index 2019 Mid Year, v0.3.7. *NASA Socioeconomic Data and Applications Center (SEDAC)*. <https://doi.org/10.7927/yzp7-sm30>. Accessed March 2023.
 56. Marin-Ferrer, M., Vernaccini, L. & Poljansek, K. Index for Risk Management INFORM Concept and Methodology Report – Version 2017. *EUR 28655 EN*. <https://doi.org/10.2760/094023> (2017).
- G.G. also contributed significantly to maps and figures. M.S. contributed to methods and formal analysis (specifically statistical analysis), writing and revisions, as well as funding acquisition. J.K. contributed to conceptualization, methodology, formal analysis, writing, revisions, and funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43247-024-01392-w>.

Correspondence and requests for materials should be addressed to Junghwan Kim.

Peer review information *Communications Earth & Environment* thanks Terence Day and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Clare Davis and Martina Grecequet. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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/>.

© The Author(s) 2024

Acknowledgements

C.A. and M.S. thank the support from National Science Foundation (#2206479). J.K. thanks the support from the Institute for Society, Culture and Environment at Virginia Tech.

Author contributions

C.A. conceptualized and proposed the study, carried out methodology and formal analysis, and wrote and revised the manuscript. G.G. contributed to conceptualization, methodology, formal analysis, writing, and revisions.