nature climate change

Perspective

https://doi.org/10.1038/s41558-023-01890-3

The next generation of machine learning for tracking adaptation texts

Received: 30 January 2023

Accepted: 13 November 2023

Published online: 27 December 2023

Check for updates

Anne J. Sietsma[1,2] ✉, James D. Ford[2] & Jan C. Minx[2,3]

Machine learning presents opportunities for tracking evidence on climate change adaptation, including text-based methods from natural language processing. In theory, such tools can analyse more data in less time, using fewer resources and with less risk of bias. However, the first generation of adaptation studies have delivered only proof of concepts. Reviewing these first studies, we argue that future efforts should focus on creating more diverse datasets, investigating concrete hypotheses, fostering collaboration and promoting 'machine learning literacy', including understanding bias. More fundamentally, machine learning enables a paradigmatic shift towards automating repetitive tasks and makes interactive 'living evidence' platforms possible. Broadly, the adaptation community is failing to prepare for this shift. Flagship projects of organizations such as the IPCC could help to lead the way.

As the climate crisis continues, the need to respond to unavoidable impacts is increasing. This is known as adaptation, that is, "[t]he process of adjustment to actual or expected climate and its effects in order to moderate harm or exploit beneficial opportunities" (Box TS.1 in ref. 1). Consequently, the importance of tracking adaptation also increases: good high-level overviews of adaptation efforts can help communities to learn from each other and direct resources to where they are most needed[2]. Yet, despite political urgency (paragraphs 74–77 in ref. 3) and a multitude of frameworks and methodologies to do this[4–7], global tracking of the effectiveness and progress of adaptation actions has proven difficult[8,9]. Current efforts tend to use proxies, including policy texts or funding flows[4,5,10], or they document the state of scientific evidence, for example, in the form of systematic maps[11–13]. Such approaches do not measure adaptation outcomes ('how much risk is being reduced?'), but they do provide insight into adaptation processes ('where and how is adaptation taking place?').

In the age of 'big literature'[14], part of the problem is the sheer volume and variety of the evidence base, for example, tens of thousands of climate change papers are published every year[15], with adaptation evidence often stemming from case studies[16]. At the same time, evidence of ongoing adaptation processes has become more widely available due to digitalization and the increasing interest in adaptation in general[12].

In addition, there is an ongoing debate on how to define adaptation in general[17,18], and adaptation success[4,8] in particular. Add to this the urgency of the climate crisis, and it becomes clear that any attempt to track adaptation progress will need to be capable of rapidly handling large and varied datasets, while simultaneously remaining sensitive to fine-grained distinctions and context-dependent meanings.

In theory, this is exactly what machine learning (ML) promises: to take human-like decisions at scale rapidly[19,20].

Recently, that promise has been increasingly put to the test in the first generation of articles that have used ML methods to assess adaptation evidence in practice[2,12,21–23]. The goal of this Perspective is to contrast these two strands of literature—the theoretical potential and the practical application of ML for adaptation tracking—as there appears to be a mismatch: while the theoretical potential paints an overwhelmingly positive image of both current and future ML, the newly emerging experiences of those who have carried out this work are more mixed.

We include ourselves in that last category, having piloted various ML methods in an adaptation context. Our expertise lies especially with natural language processing (NLP), which, as a field, analyses all sorts of text, from social media posts to policy documents. This Perspective is rooted in these personal experiences and related literature on tracking evidence through textual data; readers interested in ML applications in areas such as image processing, remote sensing and risk modelling may wish to consider additional literature as well[24–26].

[1]Public Administration and Policy Group, Wageningen University, Wageningen, The Netherlands. [2]Priestley Centre for Climate Futures, University of Leeds, Leeds, UK. [3]Mercator Research Institute on Global Commons and Climate Change, Berlin, Germany. ✉e-mail: anne.sietsma@wur.nl

https://doi.org/10.1038/s41558-023-01890-3

Overall, despite all of the concerns and criticisms voiced here, we continue to believe that ML could transform climate change research in general and adaptation research in particular. However, realizing this potential will require researchers and practitioners to be clear-eyed about the limitations of ML and to think strategically on where it can best be applied. Big literature and ML, like climate change, are here to stay. The adaptation community urgently needs to discuss how to make the most of it.

Machine learning is both cutting-edge and established

Theoretical papers on ML and adaptation tend to focus on the future potential of ML, describing it as being novel and relatively unexplored[19–21,27]. In the meantime, a first generation of application studies has emerged. A rapid review of the literature that either uses or substantially discusses the use of ML for adaptation evidence found 54 relevant papers in Web of Science and Scopus (see Supplementary Information for the protocol and included papers). Note that this literature excludes virtually all modelling and remote-sensing work: although ML applications are gaining ground here too[24–26], these studies typically assess impacts and risks rather than adaptation. Consequently, the works discussed here largely rely on textual data. As stated, this is also where our personal expertise lies. In addition, a substantial number of additional papers discuss ML in contexts closely related to adaptation, such as vulnerability, climate change in general or sustainability. These are not included in our review, but do support the notion that ML approaches are gaining traction.

All of the included papers were published in or after 2015, and 40 of these contain primary research; the remainder describe theory or are literature reviews. We summarize the findings of a few illustrative studies in Table 1.

Although many different ML methods exist, the extant literature mostly describes a fairly small subset of methods, in particular, topic models and other clustering algorithms[13,23,28–35]. Topic models are used to create an overview of a collection of texts by identifying and quantifying topics, that is, groups of words that occur frequently together in a subset of the documents. Common topic models, such as Latent Dirichlet Allocation (LDA), are over two decades old and extremely widely cited[36].

Such unsupervised ML seeks patterns in the input data without needing any kind of hand-labelled data. This means that they are more or less ‘plug and play’: find a dataset, run the model and you will get results fairly quickly.

Of course, gathering and preparing data could still be time-intensive, for example, when the data come from survey responses[35,37]. In general though, adaptation researchers opt for existing datasets, such as self-reported data from cities[30,31], or for data that is relatively easy to obtain in a structured manner, such as UNFCCC documents[21,23,34] and especially scientific literature[12,13,29,38]. Note that in all these cases, text-based data are used as proxies for adaptation processes; the evaluation of outcomes is generally not possible based on these data. Further, these datasets tend to have poor coverage in the Global South, even though adaptation needs here are generally high[12,13,23].

Supervised ML, by contrast, is less commonly used for adaptation. These types of method ‘learn’ from a so-called training set with labelled data. For example, human coders can screen scientific or policy documents to see whether they deal with ‘adaptation’ or not; the ML model then learns from these examples to select adaptation documents from a much larger unseen text corpus[2,12,13,22]. By contrast, if the same body of text is given to an unsupervised model, it will look for patterns, but there is no guarantee that the pattern it finds distinguishes between adaptation and non-adaptation. Supervised methods therefore have a clear advantage: they can be trained to perform a specific predetermined task. The disadvantage, however, is equally clear: labelled data are rare and producing the required labels can be costly.

Table 1 | Examples of studies using ML in an adaptation context

Reference	Dataset	ML method	Sample findings
2	Primary research articles indexed in Web of Science, Scopus or Medline	Supervised learning to select and categorize implemented adaptation projects; pretrained algorithm to extract geographic locations	The number of adaptation projects is growing quickly, but research gaps persist: evidence is largely local and fragmented; evidence on transformational adaptation is limited
23	UNFCCC National Communications	Structural Topic Model (unsupervised learning)	The emphasis is on climate impacts, but shifting towards adaptation, governance and vulnerability; substantial Global North–South differences
28	Press releases of 82 cities in the United States	Support vector machine (supervised learning) to select climate-relevant texts; content analysis using seeded LDA (unsupervised learning with some user input)	The salience of adaptation is increasing overall; cities that are especially vulnerable discuss adaptation more, including some cities with Republican mayors
21	Speeches at COPs, council minutes from 25 municipalities in Canada	LDA (unsupervised learning)	A disconnect in the discourse: the Global South focuses on adaptation planning and feasibility, the Global North on finance and overlaps with mitigation; municipalities focus on extreme events and the built environment
29	Hand-selected primary research articles on human mobility and the environment in Scopus	LDA (unsupervised learning) with a clustering algorithm (not ML)	The literature is diverse; the adaptation and impact literature are relatively separate; the focus is on sudden hazards over long-term climate change

The full list of articles can be found in Supplementary Information. COP, Conference of the Parties, the main United Nations forum for climate change.

Still, some adaptation-relevant papers have reported supervised methods[2,12,13,22,39,40], and here again we find that these projects tend to rely on relatively well-established implementations, including support vector machines[2,12,28,41] and neural nets[22,39,40]. The specific workings of these widely used models are discussed elsewhere[42,43] as both have a wide variety of use cases, but in the extant NLP adaptation literature, they are typically used to categorize texts in a supervised manner. There are only a few examples of large language models (LLMs) or transformer models used for adaptation[11,44,45]. The relatively small transformer-based ClimateBERT is a noteworthy recent addition here. Such models have been trained on large text corpora to gain a relatively detailed general understanding of language, which in turn allows them to perform well on a variety of NLP tasks through so-called

Table 2 | Comparison of the theoretical promises of ML against the practical findings of projects that use ML to analyse adaptation evidence from text data, with some suggestions on how to move forward

Promise	Practice	Progress
Scale: ML methods can analyse more (diverse) data in less time	Time savings possible, but data availability and heterogeneity frequently a limitation	Prioritize projects using new data sources; establish and share systematically collected datasets
Efficiency: ML approaches require fewer resources, including less expertise, for complex assessments	Technical- and subject-specific expertise required, sometimes in the same person; current lack leads to bad science	Collaboration within universities and flagship projects; provide training on basics; actively develop and require standards
Discovery: ML methods are value-free tools that can provide unbiased novel insights	Biases in data remain, bias in models harder to counteract; models are good at creating general overviews, but not at critical assessments	Combine multiple datasets in one project; at the outset of the project, set clear goals and hypotheses

Each of the rows is elaborated on in the main text.

transfer learning[46]. There are also pretrained models[12,13,30,34,47], that is, ML algorithms that have already been trained on a different dataset for a specific task. Adaptation scholars generally use these for the relatively well-known tasks of sentiment analysis (that is, identifying what emotion is associated with a statement) or identifying geographic locations.

In sum, based on this scoping review, the prevailing image is that ML applications for adaptation tracking have so far mostly provided a first proof of concept using established methods and data. This is understandable: when trying something new, it makes sense to start with well-documented approaches. However, given that we have found a few dozen examples of recent ML projects in adaptation, we argue that broadly, the concept has already been shown to work, or at least, has shown enough promise to warrant further development.

The next two steps are important for the field to make a transition from the first generation of applications to a more mature use of ML for adaptation tracking: (1) to learn from best practices and common pitfalls, where the literature is relatively mature, and (2) to reflect on both the strategic priorities field and the opportunities afforded by ML methods that are still emerging. We will discuss each in turn.

Promise meets practice

In the following, we take three of the most oft-repeated promises of ML in the adaptation tracking literature and provide critical reflections, as well as ideas on how to make progress. We summarize each point in Table 2.

Time savings are possible but data are a bottleneck

The most-cited reason to use ML for adaptation tracking is its ability to assess more data in less time. This is an exciting promise, and the good news is that ML often manages to deliver on this in practice.

Supervised methods have shown good results for literature reviews in particular, and these need not require any programming skills: there are multiple off-the-shelf platforms that use NLP methods to prioritize documents that are likely to be relevant[48,49]. This can cut the review time in half or less[48,50]. Such approaches are especially useful for searches that return a few thousand results, of which perhaps a few hundred are relevant, meaning that after the initial screening, detailed analysis can still take place by hand. These kinds of numbers are common for reviews of subtopics within adaptation[51–53].

For even larger subjects, it may be better to train a new ML model[48]. This requires additional knowledge and time to set up and annotate the training data. Based on our personal experience, training a supervised model to select relevant abstracts of scientific papers often requires a few hundred positive examples, depending on the complexity of the task, which, in our experience, often translates to 2,000–4,000 screened articles. This implies a substantial amount of time labelling articles—even if one article takes 1 minute to label, that amounts to around 50 hours—but if the complete search returned, for example, 15,000 documents, manual screening of all abstracts would take roughly 5 times longer still. Larger searches may benefit even more[11,12], although care should be taken as this also increases the risk that some areas of the literature are not sufficiently present in the training sample, which would lead them to be under-represented in the final outcome.

The picture changes somewhat when we consider unsupervised NLP methods such as topic modelling or word embeddings from LLMs. Because there is no need for a labelled training set, the most time-consuming component of many supervised approaches is removed. This means that the time investment is broadly similar to, for example, bibliometric analyses; however, those rely on relatively crude heuristics (for example, keywords or the number of times a single word is used), whereas methods such as topic modelling can provide more granular insights into the content of a document set. Note, conversely, that meaningful validation of unsupervised methods can be complex and time-intensive, requiring a mixture of statistical and quantitative methods[54], a step that is often marginalized in practice. Overall, unsupervised ML is not necessarily quicker than established alternatives, but it can provide detailed insights relatively easily, making it well suited to exploratory analyses and tracking trends in larger datasets where more qualitative analyses are no longer feasible.

This begs the question, what is an appropriate size for a dataset? The lower limit depends on both the model and the task. When tracking evidence from text data, this limit is most likely to be a concern for specialist topics: the model probably needs more examples to 'learn' to make the required fine-grained distinctions, but at the same time, these specialist examples are rare and therefore difficult to find. In addition, for such smaller datasets, manual analysis is usually feasible and will provide more detailed insights, so the added value of ML is negligible. In our experience, for document analysis, ML tools therefore are useful if there are at least a few hundred documents on the subject of interest[11–13]. Note, however, that this threshold may be lowered in the near future; LLMs in particular are getting better at a wide range of tasks, including zero-shot learning[55] and generating synthetic data[56]. More generally, their 'emergent abilities'[57] mean that they are capable of performing tasks that they were never formally trained for at all. While promising, this also highlights the urgent need for evaluating the performance and trustworthiness of such models.

The upper limit on dataset size is even less clear. One limiting factor may be computing power; especially when using LLMs such as ChatGTP or Bard, or when training transformer models such as BERT and its successors, computer clusters with generous amounts of memory and graphics cards may be required. Less well-resourced projects therefore may reasonably consider whether the improvements in classifier performance are worth this cost. Still, the wide availability of cloud computing platforms and application programming interfaces means that the size of the dataset is rarely, if ever, the main limitation for well-resourced NLP projects.

Instead, the upper limit is often set by data availability and heterogeneity. As noted earlier, so far, adaptation tracking literature has tended to focus on well-established data sources. A common suggestion is that future research should include more diverse sources, especially the so-called grey literature. However, combining different datasets or manually annotating data is time-intensive, and grey literature in particular is difficult to work with: the Global Adaptation Mapping Initiative[2] relied on a large team of 126 researchers, but even this proved insufficient to systematically include grey literature. Relatedly, Hsu and Rauber[31] provide one of the few examples where a substantial number of databases are combined, but they still caution that their data largely originate from Europe and "is limited by the lack of time-series data, regular and repeated reporting on climate actions, strategies, and policies". In other words, rather than analysing 'more data in less time', often, NLP projects analyse 'more of the same data in less time' because different data might not exist or are too difficult to retrieve systematically.

This is not to say that including alternative sources should not be done, but rather that it will take considerably more effort in the absence of standardized databases[58] and methods. Researchers could, for example, use web scrapers to specifically target government websites in areas where traditional data coverage is poor (for example, many areas in the Global South). Combining different sources will require additional experimentation, for example, with automated summarizers to create document sets of a more homogeneous length, by translating non-English data automatically or by using multilanguage models. For adaptation tracking in scientific texts, we see a large role for database providers and libraries[59] who could more systematically index non-academic sources and make them available in a standardized computer-readable format. This would broaden the scope of reviews in general, as well as making it easier to leverage the time savings and broader scope of ML-assisted reviews.

Topical expertise and ML literacy both needed

A second commonly cited promise of ML approaches is that they can efficiently handle complex data. Because ML systems can adapt to a wide variety of inputs and can learn to make relatively granular distinctions without explicitly being programmed to do so, the implication is that ML approaches require smaller teams who need to spend less time becoming a topical expert as 'the computer' in many ways does the heavy lifting. In practice, however, not only is this untrue but also it can lead to bad science, including poorly designed or executed research and problems with peer review.

Even if technical skill could substitute topical expertise, these technical skills are often lacking in the adaptation community[19,21]. Platforms and well-designed tools may lower this barrier to entry, and the difficulty of writing computer code itself may also decrease as ML models become better at translating plain language instructions into code, although it may be some time yet before this is sufficiently reliable[60]. These positive developments notwithstanding, some technical expertise is always required. Without it, researchers may have unrealistic expectations of what the ML system can achieve, or they might over-interpret the results.

A lack of technical expertise also affects the peer-review process for projects using ML. Consider, for example, performance scores for classifiers: the easiest option is simple accuracy, meaning the percentage of correct classifications; however, if only 10% of documents are relevant, a (practically useless) classifier can still have an accuracy of 90% by predicting that all documents are irrelevant. Computer scientists therefore commonly report an F1 score instead, which compensates for unbalanced data. The F1 score is typically around 70–90% for binary problems[11,12,15,22], but it may be much lower for complex tasks[44,61]. Unless reviewers have a background in ML, they will probably have no appropriate frame of reference to evaluate whether a given score is reasonable for the problem at hand. As a result, researchers may report the accuracy or other well-known statistics anyway[39,62], or place performance scores in the supplementary material[2,45,63], which avoids technical explanations and questions from reviewers but makes results more difficult to interpret. A broader community with technical expertise would avoid this.

Further, in our experience, topical expertise remains as important as in traditional research set-ups. Without it, researchers will not be able to ask the right questions, or operationalize and execute projects adequately. To give an example in NLP, consider how documents on adaptation might be found. Given that a large dataset is less of a concern for NLP methods, one might opt for a simple query using general terms, combining different forms of 'climate' and 'adaptation' (for example, climat* AND adapt*). ML methods could then be used to remove the irrelevant results afterwards. However, relying only on general terms can give a false sense of completeness. The above example would miss many relevant articles, including from the disaster risk reduction literature, as the climate component of many natural disasters is not always explicitly named in the abstract; a researcher may even want to include keywords for mitigation (for example, mitigat*) in the search as 'risk mitigation' is sometimes used instead of 'adaptation'[51,64]. The easiest, and arguably least visible, way of introducing bias is by leaving out data that you did not know existed. Domain-specific knowledge is required to cover such blind spots.

A similar dynamic plays out when analysing the results. Take, for example, the outcomes of a topic model. Although these models 'discover' topics, this does not mean that the background knowledge needed to construct viable topics is obsolete, as topic models require knowledge of the subject to interpret[21]. There are two caveats here: first, some quantitative measures for topic model quality do exist[65] and second, topic models are sometimes used purely to explore data, in which case it is more defensible to have limited a priori knowledge. For most analyses though, including scientific research, results need to be contextualized and critically analysed, which requires domain-specific expertise.

Collaborations between computer scientists and domain experts may help to bring the required knowledge together. For academia, the Climate Change AI community[27] has set up climate change tracks at computer science conferences. Conversely, we would urge the organizers of adaptation conferences to also actively invite the ML community (for example, Adaptation Futures or European Climate Change Adaptation). Universities and individual academics can foster interdisciplinary collaborations too; for reasons of space, we point to the recent overview by Lyall[66] for this much broader topic.

Still, in our personal experience, it is not always enough to simply create a team with a domain expert and a topic expert. Interdisciplinary research broadly recognizes that combining different domain-specific epistemologies is often difficult and time-consuming[67,68]. In ML projects specifically, the model parameters will influence the outcome and are dependent on the data. This means that a deep understanding of both the methods and the data is required, first to select the appropriate methods, as well as to distinguish between methodological artefacts and meaningful results. In other words, we find that topical and domain knowledge are at times required in the same person.

Ultimately, what is needed is an active community of practice. Training would, in theory, help to create this; however, training materials on ML, including NLP, have been widely available for some time, yet adaptation applications are few and far between. We therefore believe that adaptation organizations should focus first on improving 'machine learning literacy', that is, consciously aiming for breadth over depth so that a wider subsection of the community will have a basic understanding of the central concepts and methods. This would help adaptation practitioners to recognize opportunities for ML in their own work, while also ensuring that results can be fruitfully discussed and that researchers can critically reflect on possible improvements and next steps. This, in turn, could feed into concrete guidelines and best practices for ML in climate change research. In our view, open science and systematic review standards[69] should be the starting point of such guidelines, both to increase transparency and to accelerate progress. In sum, focusing on the basics can help to create space for a new generation of researchers to develop shared goals and norms.

Models repeat biases and conventional wisdoms

The third promise that we wish to examine is perhaps best exemplified by the creators of the Structural Topic Model, who state that a topic model "allows the researcher to *discover* topics from the data,

https://doi.org/10.1038/s41558-023-01890-3

rather than assume them" (page 1,066 in ref. 70). This quote and similar formulations are used to make two closely related points: first, it suggests that topic models are less biased[21,29,31] and second, that these tools would lead to new insights as they can "uncover hidden patterns" (page 136 in ref. 71) and "identify facts and relationships that would otherwise remain buried" (page 4 in ref. 47). Such sweeping claims deserve scrutiny.

Strictly speaking, it is true that computers simply 'do as they are told', but this does not necessarily equate to less bias; rather, computational methods have their own biases[72]. Running a topic model on biased data, for example, will still give biased results. An argument could even be made that, by treating all the data as equally valuable, topic models are less suited than more critical qualitative methods to deal with unbalanced datasets. Equally, we have ourselves found that topic models can be useful for identifying quantitative gaps in adaptation evidence[2,13,44], but this requires the researcher to know the field well enough to see which topics should be in the outcome but are not.

For supervised methods, the model will replicate the bias of the people who created the training data, for example, there is an ongoing and politically charged debate on what differentiates adaptation from general development[17], so if one is trying to teach a supervised classifier to make this distinction, the personal beliefs of an individual reviewer may well influence their judgements. The remedy here is the same as with a traditional review: publish a clear protocol, preferably with a diverse group of stakeholders.

However, this is impossible to do with pretrained models. There are a variety of methods to quantify or adjust for biases in these models[73,74], but to some degree, one simply has to trust the original creators. This is especially true for LLMs, as training these models takes enormous amounts of resources[75]. Meanwhile, the bias of LLMs around gender, race and religion, among others, is well documented[76,77]. How much this affects adaptation specifically has not been studied systematically; doing so is beyond the scope of this Perspective, but we give some examples in Table 3. As climate impacts, vulnerabilities and adaptive capacities are intertwined with broader issues of justice and inequality, such bias can be highly problematic.

To be clear, we do not mean to imply that ML methods are always inherently flawed, but where scientists have over the years built up a considerable arsenal of methods to account for bias in traditional methods, these tools are still very much under development for ML approaches.

Similarly, ML methods can certainly be used to generate novel insights, but it is 'data hubris'[78] to think that such insights will simply reveal themselves with sufficient data and algorithms. Even the most cutting-edge NLP models have been called 'stochastic parrots'[79] that lack critical thinking and cannot distinguish between what is widely stated and what is widely (dis)proven. Further, in the rare cases where ML outcomes are compared with expert opinions (for example, refs. 12,30), the model is more likely to agree with expert opinion than to lead to fundamentally new understandings.

Overall, if done right, ML can be a useful tool to produce new knowledge, but using a novel method by itself does not guarantee novel outcomes. In our experience, there are two main ingredients that increase the likelihood of original insights. First, using unconventional or understudied sources of data and layering different data sources makes it easier to identify diverging patterns, for example, comparing twitter discourse and academic publishing[80], and overlaying grid-cell climate models and observations with the location and topics of climate impact studies[11]. As noted earlier, however, heterogeneous datasets will take considerable effort to create.

Second, it helps to go in with a clear notion of what is expected or desired. Obviously, 'fishing expeditions' should be avoided, but an uncritical analysis is virtually guaranteed to result in only well-known broad trends. Examples include looking for shifts in national reporting post-Paris Agreement and finding that such shifts are barely perceptible[81], and investigating whether the intended interlinkages between different Sustainable Development Goals are found in practice[82]. In this light, it is worth noting that formal hypothesis testing and error ranges are often not reported in ML-assisted syntheses, with the reports in refs. 11,12 being exceptions.

Table 3 | Examples of potential bias using the ClimateBERT language model to fill in the blank

Prompt	Most likely words	Bias
Climate change adaptation [blank] women	for (34.5%) by (13.6%)	Women are seen as victims who are recipients of adaptation efforts rather than actors with agency[91,92]
Climate change adaptation [blank] men	by (27.7%) for (23.3%)	
Adaptation in the USA is [blank]	underway (15.0%) ongoing (9.0%)	The focus in Bangladesh is on the vulnerability and the need for more action, while the USA is depicted as a place where adaptation is already happening
Adaptation in Bangladesh is [blank]	critical (10.9%) urgent (10.5%)	
Effective adaptation requires [blank]	partnerships (17.6%) innovation (17.2%)	Adaptation is seen as a local effort in vulnerable places where people need to work together to overcome climate risks, while mitigation is something that individuals can start doing
Effective mitigation requires [blank]	action (16.6%) innovation (14.8%)	
Ali [blank] climate change	denies (9.1%) blamed (6.6%)	A common name in predominantly Muslim countries is associated with negative terms and climate denial, whereas a common name in English-speaking countries results in neutral words
Smith [blank] climate change	on (11.1%) discussed (7.2%)	
The task was given to the project leader; [blank] completed it	he (49.1%) they (21.0%)	Project leaders are assumed to be men more often than other genders ('she' scored a probability of 2.5% in the first example, while 'they' scored 3.5% in the second)
Adaptation support was provided by the minister; [blank] visited personally	he (53.1%) she (10.3%)	
The storm made landfall in [blank]	Louisiana (36.2%) Alabama (13.4%)	The model assumes an American and Northern Hemisphere perspective, probably because a disproportionate amount of climate research originates there[12] ('September' scores a probability of 6.7% in the second example)
The summer starts in [blank]	June (13.9%) May (13.1%)	

Examples were generated using a transformer-based model. Such models are created by training on large sets of documents; here we used ClimateBERT, for which the training includes climate change documents. The model can be used, as we did here, to calculate what word is most likely to occur in a given place in a sentence (that is, 'fill in the blank'). We give the two most likely words with their assigned probabilities and explain why this can be seen as evidence of bias.

Treating ML as a transformation

So far, we have focused on the main promises of ML in existing literature. We find that the majority of this literature is concerned with fitting ML into business as usual, that is, the same report but bigger. In our opinion, however, the real revolution enabled by abundant data and computational power is not one of degree, but one of kind.

Traditional research leans on a few highly trained individuals, but ML excels at doing simpler tasks thousands of times, which opens up entirely new approaches.

To be clear, we are not arguing for the wholesale replacement of established qualitative methods. Manual and computer-based methods can and should coexist, but they will compete for resources. It is therefore worth thinking critically on the types of insight that are most urgently needed at different stages of adaptation projects, and what combination of methods and final products can create diverse, complementary and comprehensive lines of evidence. Insofar as that includes NLP, we contend that in many cases, the best approach is unlikely to be a decades-old method applied to whatever dataset happens to be easy to get. Rather, ML applications should be ambitious, build on the work that has already been done and play to the strengths of the method.

An underrated element in the ML revolution is how easily such projects can be repeated. Curating the dataset and developing the initial model is often the time-consuming part; once the code for this has been written, it is relatively straightforward to re-run the code on either a different time point or different dataset (although it may be necessary to update or fine-tune the model). This makes ML especially useful for the types of repetitive task that form the foundation of many adaptation projects, such as finding adaptation-relevant passages in policy documents[22], linking adaptation evidence to locations and impact models[11], identifying knowledge gaps[12,83] or, more ambitiously, creating a standardized global database of adaptation interventions[58]. As an aside, the training of models to perform such general tasks could be conceived as a public service, which therefore should receive public funding. This would also help to alleviate the unequal access to computational power.

In addition, if the whole 'pipeline' of an NLP-based tracking system is re-run at regular intervals, a near-real-time overview of the evidence could be created. To some, self-updating tracking systems may seem futuristic, but the reality is that, technologically, this is already feasible. These so-called living evidence approaches have recently gained popularity, especially in the health sciences[84,85]. Although living reviews may still have a manual component (for example, updating models), it is easy to foresee (semi-)automated systems to track adaptation in science, policy and society, similar to some of the platforms that emerged during the COVID-19 pandemic[86].

These platforms highlight another missing component for the optimal use of ML: interactivity. Given its context-dependent nature and reliance on case studies, a global overview of adaptation evidence is probably too general to be useful for practitioners. ML (and data science more broadly) can be used to augment such messy data with key characteristics (typical NLP examples would be the topic, geographic location and time period) that can then form the basis for an easily searchable platform and interactive graphics. Although there are recent examples of projects that incorporate some of these suggestions, such as Climate Policy Radar[87], overall, interactive platforms are rare. Scientific outputs in particular often take the form of a table or comma-separated file. Why does research rely so heavily on old standards with less functionality than the website of almost any online store?

A large part of the answer is the continued importance of traditional publishing. Interactive figures, too, are technically entirely feasible in online editions—data science blogs routinely include runnable code—but papers are mostly still considered a finished and therefore static entity. As datasets are relegated to the supplementary material, maintaining a database does not result in new publications. Regrettably, we do not see this changing any time soon, but do encourage researchers to start exploring tools for making interactive dashboards, such as Shiny apps in R. We expand on a 'gold standard' interactive and living evidence platform in Box 1.

Repeatability and interactivity are just two examples. More conceptually, we urge the adaptation community to take seriously the paradigmatic shift presented by computational methods, including ML. This is not easy: technological advances are rapid, and some techniques may have applications that are difficult to foresee. This is particularly true for text-based analyses, where the full effects of the recent LLM revolution defy prediction[88,89]. Researchers need to tread a careful line here: while LLMs can remove common bottlenecks, in the absence of robust domain-specific evaluation metrics, human-in-the-loop systems may be a more prudent path[90]. Even so, making the most of the full breadth of ML tools will require some foresight and planning, especially around identifying the types of task that ML would be best suited for. When combining this sense of purpose with both an open mind to practices from other fields and a realistic understanding of current possibilities and limitations, adaptation evidence tracking could help protect people globally from the impacts of climate change. But this is no small task; the adaptation community has work to do.

BOX 1

Why organizations such as the IPCC should create flagship projects

A pressing example of the need for novel tracking methods is provided by the IPCC, which is entering its seventh assessment period. Their mandate to synthesize all available climate research is increasingly difficult to meet: the latest Working Group II report relies on hundreds of authors and includes over 34,000 references[1], yet despite this mammoth effort, the number of relevant articles is several times larger still and the share of this wider literature that the IPCC can incorporate is decreasing[11,93,94]. Similarly, because the reports are only issued every few years, they can lag behind the research frontier on key emerging issues[12].

Building on much older critiques[95,96], some have recently argued that the science is clear, that the IPCC has therefore served its purpose and now should be transformed into a more agile entity that produces targeted reports[97,98] (for example, on policy implementation[99]). In our view, it is clear that the IPCC at least needs to innovate, and the organization would be especially well placed to maintain an interactive living evidence platform.

The practical advantages are clear: more timely and more transparent overviews of evidence where users can easily tailor the information to their own needs. As an example, a policymaker could use such a platform to select evidence on the topic of coastal flood defences, filter for documents from government websites, select only those from tropical countries and sort the result by publication date to quickly find the most relevant passages for their work.

At the community level, such a project could have positive knock-on effects as it would help to establish a 'gold standard' for ML work in climate change. An ambitious and well-maintained platform can showcase what is currently possible, which helps attract additional talent. Over time, this will pay dividends.

There are other organizations that could also play such a role, for example, the United Nations Environment Programme, which writes the Adaptation Gap Reports. Depending on political circumstances, this may be more expedient, although the IPCC is a trusted source, which would help to ensure that the platform is used in practice. Either way, funding bodies, and in the case of the IPCC, also national governments, hold the key to unlocking this potential.

Data availability

All of the data generated or analysed during this study are included in this published article and its supplementary information files.

References

1. IPCC *Climate Change 2022: Impacts, Adaptation, and Vulnerability* (eds Pörtner, H.-O. et al.) (Cambridge Univ. Press, 2022).
2. Berrang-Ford, L. et al. A systematic global stocktake of evidence on human adaptation to climate change. *Nat. Clim. Change* **11**, 989–1000 (2021).
3. *Sharm el-Sheikh Implementation Plan* FCCC/PA/CMA/2022/L.21 (UNFCCC, 2022).
4. Berrang-Ford, L. et al. Tracking global climate change adaptation among governments. *Nat. Clim. Change* **9**, 440–449 (2019).
5. Craft, B. & Fisher, S. Measuring the adaptation goal in the global stocktake of the Paris Agreement. *Clim. Policy* **18**, 1203–1209 (2018).
6. Njuguna, L., Biesbroek, R., Crane, T. A., Tamás, P. & Dewulf, A. Designing fit-for-context climate change adaptation tracking: towards a framework for analyzing the institutional structures of knowledge production and use. *Clim. Risk Manag.* **35**, 100401 (2022).
7. Olhoff, A., Väänänen, E. & Dickson, B. in *Resilience* (eds Zommers, Z. & Alverson, K.) 51–61 (Elsevier, 2018).
8. Dilling, L. et al. Is adaptation success a flawed concept? *Nat. Clim. Change* **9**, 572–574 (2019).
9. Leiter, T. & Pringle, P. in *Adaptation Metrics: Perspectives on Measuring, Aggregating and Comparing Adaptation Results* (eds Christiansen, L. et al.) 29–48 (UNEP DTU, 2018).
10. *Adaptation Gap Report 2022: Too Little, Too Slow—Climate Adaptation Failure Puts World at Risk* (UNEP, 2022).
11. Callaghan, M. et al. Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nat. Clim. Change* **11**, 966–972 (2021).
12. Sietsma, A. J., Ford, J. D., Callaghan, M. W. & Minx, J. C. Progress in climate change adaptation research. *Environ. Res. Lett.* **16**, 054038 (2021).
13. Berrang-Ford, L. et al. Systematic mapping of global research on climate and health: a machine learning review. *Lancet Planet. Health* **5**, e514–e525 (2021).
14. Nunez-Mir, G. C., Iannone, B. V.III, Pijanowski, B. C., Kong, N. & Fei, S. Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods Ecol. Evol.* **7**, 1262–1272 (2016).
15. Callaghan, M. W., Minx, J. C. & Forster, P. M. A topography of climate change research. *Nat. Clim. Change* **10**, 118–123 (2020).
16. Nalau, J. & Verrall, B. Mapping the evolution and current trends in climate change adaptation science. *Clim. Risk Manag.* https://doi.org/10.1016/j.crm.2021.100290 (2021).
17. Schipper, E. L. F., Tanner, T., Dube, O. P., Adams, K. M. & Huq, S. The debate: is global development adapting to climate change? *World Dev. Perspect.* **18**, 100205 (2020).
18. Siders, A. R. Adaptive capacity to climate change: a synthesis of concepts, methods, and findings in a fragmented field. *Wiley Interdiscip. Rev. Clim. Change* **10**, e573 (2019).
19. Ford, J. D. et al. Big data has big potential for applications to climate change adaptation. *Proc. Natl Acad. Sci. USA* **113**, 10729–10732 (2016).
20. Cheong, S. M., Sankaran, K. & Bastani, H. Artificial intelligence for climate change adaptation. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **13**, e1459 (2022).
21. Lesnikowski, A. et al. Frontiers in data analytics for adaptation research: topic modeling. *Wiley Interdiscip. Rev. Clim. Change* **10**, e576 (2019).
22. Biesbroek, R., Badloe, S. & Athanasiadis, I. N. Machine learning for research on climate change adaptation policy integration: an exploratory UK case study. *Reg. Environ. Change* **20**, 85 (2020).
23. Biesbroek, R., Wright, S. J., Eguren, S. K., Bonotto, A. & Athanasiadis, I. N. Policy attention to climate change impacts, adaptation and vulnerability: a global assessment of National Communications (1994–2019). *Clim. Policy* **22**, 97–111 (2022).
24. Zennaro, F. et al. Exploring machine learning potential for climate change risk assessment. *Earth Sci. Rev.* **220**, 103752 (2021).
25. Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A. & Kumar, V. Machine learning for the geosciences: challenges and opportunities. *IEEE Trans. Knowl. Data Eng.* **31**, 1544–1554 (2018).
26. Munawar, H. S., Hammad, A. W. A. & Waller, S. T. A review on flood management technologies related to image processing and machine learning. *Autom. Constr.* **132**, 103916 (2021).
27. Rolnick, D. et al. Tackling climate change with machine learning. *ACM Comput. Surv*. **55**, 42 (2022).
28. Boussalis, C., Coan, T. G. & Holman, M. R. Communicating climate mitigation and adaptation efforts in American cities. *Climate* **7**, 45 (2019).
29. Zander, K. K. et al. Topic modelling exposes disciplinary divergence in research on the nexus between human mobility and the environment. *Humanit. Soc. Sci. Commun.* **9**, 34 (2022).
30. Fu, X., Li, C. & Zhai, W. Using natural language processing to read plans. *J. Am. Plann. Assoc.* https://doi.org/10.1080/01944363.2022.2038659 (2022).
31. Hsu, A. & Rauber, R. Diverse climate actors show limited coordination in a large-scale text analysis of strategy documents. *Commun. Earth Environ.* **2**, 30 (2021).
32. Abarca-Alvarez, F. J., Navarro-Ligero, M. L., Valenzuela-Montes, L. M. & Campos-Sánchez, F. S. European strategies for adaptation to climate change with the Mayors Adapt initiative by self-organizing maps. *Appl. Sci.* **9**, 3859 (2019).
33. Paulvannan Kanmani, A., Obringer, R., Rachunok, B. & Nateghi, R. Assessing global environmental sustainability via an unsupervised clustering framework. *Sustainability* **12**, 563 (2020).
34. Valero, S. D., Emandi, R., Encarnacion, J., Kaul, S. & Seck, P. Utilizing big data to measure key connections between gender and climate change 1. *Stat. J. IAOS* **38**, 973–994 (2022).
35. Lynam, T. Exploring social representations of adapting to climate change using topic modeling and Bayesian networks. *Ecol. Soc.* https://doi.org/10.5751/ES-08778-210416 (2016).
36. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
37. Tvinnereim, E., Fløttum, K., Gjerstad, Ø., Johannesson, M. P. & Nordø, Å. D. Citizens' preferences for tackling climate change. Quantitative and qualitative analyses of their freely formulated solutions. *Glob. Environ. Change* **46**, 34–41 (2017).
38. Sun, Y.-L., Zhang, C.-H., Lian, Y.-J. & Zhao, J.-M. Exploring the global research trends of cities and climate change based on a bibliometric analysis. *Sustainability* **14**, 12302 (2022).
39. Rana, I. A., Lodhi, R. H., Zia, A., Jamshed, A. & Nawaz, A. Three-step neural network approach for predicting monsoon flood preparedness and adaptation: application in urban communities of Lahore, Pakistan. *Urban Clim.* **45**, 101266 (2022).
40. Canon, M. J., Satuito, A. & Sy, C. Determining disaster risk management priorities through a neural network-based text classifier. In *2018 International Symposium on Computer, Consumer and Control (IS3C)* 237–241 (2018).
41. Salam, R. et al. Nexus between vulnerability and adaptive capacity of drought-prone rural households in northern Bangladesh. *Nat. Hazards* **106**, 509–527 (2021).
42. Abiodun, O. I. et al. State-of-the-art in artificial neural network applications: a survey. *Heliyon* **4**, e00938 (2018).
43. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. & Lopez, A. A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing* **408**, 189–215 (2020).
44. Sietsma, A. J. et al. Machine learning evidence map reveals global differences in adaptation action. *One Earth* (in the press).

45. Bingler, J. A., Kraus, M., Leippold, M. & Webersinke, N. Cheap talk and cherry-picking: what ClimateBERT has to say on corporate climate risk disclosures. *Financ. Res. Lett.* **47**, 102776 (2022).
46. Gillioz, A., Casas, J., Mugellini, E. & Khaled, O. A. Overview of the transformer-based models for NLP Tasks. In *2020 15th Conference on Computer Science and Information Systems (FedCSIS)* 179–183 (2020).
47. Huo, F. et al. Using big data analytics to synthesize research domains and identify emerging fields in urban climatology. *WIREs Clim. Change* **12**, e688 (2021).
48. Khalil, H., Ameen, D. & Zarnegar, A. Tools to support the automation of systematic reviews: a scoping review. *J. Clin. Epidemiol.* **144**, 22–42 (2022).
49. Marshall, C., Sutton, A., O'Keefe, H. & Johnson, E. (eds) *The Systematic Review Toolbox* (2022); http://www.systematicreviewtools.com/
50. Gates, A. et al. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. *Syst. Rev.* **8**, 278 (2019).
51. Bisaro, A., Roggero, M. & Villamayor-Tomas, S. Institutional analysis in climate change adaptation research: a systematic literature review. *Ecol. Econ.* **151**, 34–43 (2018).
52. Scheelbeek, P. F. et al. The effects on public health of climate change adaptation responses: a systematic review of evidence from low- and middle-income countries. *Environ. Res. Lett.* **16**, 073001 (2021).
53. Naulleau, A., Gary, C., Prévot, L. & Hossard, L. Evaluating strategies for adaptation to climate change in grapevine production—a systematic review. *Front. Plant Sci.* **11**, 607859 (2021).
54. Müller-Hansen, F., Callaghan, M. W. & Minx, J. C. Text as big data: develop codes of practice for rigorous computational text analysis in energy social science. *Energy Res. Soc. Sci.* **70**, 101691 (2020).
55. Wu, P. Y., Tucker, J. A., Nagler, J. & Messing, S. Large language models can be used to estimate the ideologies of politicians. Preprint at https://arxiv.org/abs/2303.12057 (2023).
56. Hämäläinen, P., Tavast, M. & Kunnari, A. Evaluating large language models in generating synthetic HCI research data: a case study. In *Proc. 2023 CHI Conference on Human Factors in Computing Systems* 1–19 (2023).
57. Wei, J. et al. Emergent abilities of large language models. Preprint at https://arxiv.org/abs/2206.07682 (2022).
58. Canales, N., Klein, R. J. T., Bakhtaoui, I. & Macura, B. Assessing adaptation progress for the global stocktake. *Nat. Clim. Change* https://doi.org/10.1038/s41558-023-01656-x (2023).
59. Marsolek, W., Farrell, S. L., Kelly, J. A. & Cooper, K. Grey literature: advocating for diverse voices, increased use, improved access, and preservation. *Coll. Res. Libr. News* https://doi.org/10.5860/crln.82.2.58 (2021).
60. Poldrack, R. A., Lu, T. & Beguš, G. AI-assisted coding: experiments with GPT-4. Preprint at https://arxiv.org/abs/2304.13187 (2023).
61. Corringham, T. et al. BERT classification of Paris Agreement climate action plans. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning* 45 (2021).
62. Manandhar, A. et al. Machine learning to evaluate impacts of flood protection in Bangladesh, 1983–2014. *Water* **12**, 483 (2020).
63. Sachdeva, S., Hsu, A., French, I. & Lim, E. A computational approach to analyzing climate strategies of cities pledging net zero. *npj Urban Sustain.* **2**, 21 (2022).
64. Kim, B. J., Jeong, S. & Chung, J.-B. Research trends in vulnerability studies from 2000 to 2019: findings from a bibliometric analysis. *Int. J. Disaster Risk Reduct.* **56**, 102141 (2021).
65. Jacobs, T. & Tschötschel, R. Topic models meet discourse analysis: a quantitative tool for a qualitative approach. *Int. J. Soc. Res. Methodol.* **22**, 469–485 (2019).
66. Lyall, C. *Being an Interdisciplinary Academic—How Institutions Shape University Careers* (Springer Nature, 2019).
67. MacLeod, M. What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. *Synthese* **195**, 697–720 (2018).
68. Miller, T. R. et al. Epistemological pluralism reorganizing interdisciplinary research. *Ecol. Soc.* **13**, 46 (2008).
69. Haddaway, N. R., Macura, B., Whaley, P. & Pullin, A. S. ROSES Reporting Standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environ. Evid.* **7**, 7 (2018).
70. Roberts, M. E. et al. Structural topic models for open-ended survey responses. *Am. J. Pol. Sci.* **58**, 1064–1082 (2014).
71. Miglionico, A. The use of technology in corporate management and reporting of climate-related risks. *Eur. Bus. Organ. Law Rev.* **23**, 125–141 (2022).
72. Hovy, D. & Prabhumoye, S. Five sources of bias in natural language processing. *Lang. Linguist. Compass* **15**, e12432 (2021).
73. Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183–186 (2017).
74. Guo, Y., Yang, Y. & Abbasi, A. Auto-debias: debiasing masked language models with automated biased prompts. In *Proc. 60th Annual Meeting of the Association for Computational Linguistics* 1012–1023 (ACL, 2022).
75. Xu, F. F., Alon, U., Neubig, G. & Hellendoorn, V. J. A systematic evaluation of large language models of code. In *Proc. 6th ACM SIGPLAN International Symposium on Machine Programming* 1–10 (ACM, 2022).
76. Garrido-Muñoz, I., Montejo-Ráez, A., Martínez-Santiago, F. & Ureña-López, L. A. A survey on bias in deep NLP. *Appl. Sci.* **11**, 3184 (2021).
77. Magee, L., Ghahremanlou, L., Soldatic, K. & Robertson, S. Intersectional bias in causal language models. Preprint at https://arxiv.org/abs/2107.07691 (2021).
78. Lazer, D., Kennedy, R., King, G. & Vespignani, A. The parable of Google Flu: traps in big data analysis. *Science* **343**, 1203–1205 (2014).
79. Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. On the dangers of stochastic parrots: can language models be too big? In *Proc. 2021 ACM Conference on Fairness, Accountability, and Transparency* 610–623 (ACM, 2021).
80. Haunschild, R., Leydesdorff, L., Bornmann, L., Hellsten, I. & Marx, W. Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags. *J. Informetr.* **13**, 695–707 (2019).
81. Wright, S. J., Sietsma, A. J., Korswagen, S., Athanasiadis, I. N. & Biesbroek, R. How do countries frame climate change? A global comparison of adaptation and mitigation in UNFCCC National Communications. *Reg. Environ. Change* **23**, 129 (2023).
82. Smith, T. B., Vacca, R., Mantegazza, L. & Capua, I. Natural language processing and network analysis provide novel insights on policy and scientific discourse around Sustainable Development Goals. *Sci. Rep.* **11**, 22427 (2021).
83. Berrang-Ford, L. et al. in *Mapping Climate–Health Evidence: Using Machine-Learning to Map the Links Between Climate Change and Health* 192 (Foreign, Commonwealth and Development Office, 2021).
84. Elliott, J. H. et al. Living systematic review: 1. Introduction—the why, what, when, and how. *J. Clin. Epidemiol.* **91**, 23–30 (2017).

85. Millard, T. et al. Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation. *Syst. Rev.* **8**, 325 (2019).
86. Khalil, H., Tamara, L., Rada, G. & Akl, E. A. Challenges of evidence synthesis during the 2020 COVID pandemic: a scoping review. *J. Clin. Epidemiol.* **142**, 10–18 (2022).
87. Climate Policy Radar *Climate Policy Radar App* https://app.climatepolicyradar.org/ (2023).
88. Liu, Y. et al. Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology* **1**, 100017 (2023).
89. Floridi, L. & Chiriatti, M. GPT-3: its nature, scope, limits, and consequences. *Minds Mach.* **30**, 681–694 (2020).
90. Debnath, R., Creutzig, F., Sovacool, B. K. & Shuckburgh, E. Harnessing human and machine intelligence for planetary-level climate action. *npj Clim. Action* **2**, 20 (2023).
91. Huyer, S. & Gumucio, T. Going back to the well: women, agency, and climate adaptation. *World J. Agric. Soil Sci.* **5**, WJASS.MS.ID.000611 (2020).
92. Wester, M. & Lama, P. D. in *Climate Hazards, Disasters, and Gender Ramifications* (eds Kinnvall C. & Rydstrom H.) 67–85 (Routledge, 2019).
93. Minx, J. C., Callaghan, M., Lamb, W. F., Garard, J. & Edenhofer, O. Learning about climate change solutions in the IPCC and beyond. *Environ. Sci. Policy* **77**, 252–259 (2017).
94. Berrang-Ford, L. et al. Evidence synthesis for accelerated learning on climate solutions. *Campbell Syst. Rev.* https://doi.org/10.1002/cl2.1128 (2020).
95. Petticrew, M. & McCartney, G. Using systematic reviews to separate scientific from policy debate relevant to climate change. *Am. J. Prev. Med.* **40**, 576–578 (2011).
96. Tol, R. S. J. Regulating knowledge monopolies: the case of the IPCC. *Climatic Change* **108**, 827 (2011).
97. Provost, G. Rigorous and relevant: applying lessons from the history of IPCC special reports to the post-Paris Agreement world. *Harvard Environ. Law Rev.* **43**, 507–546 (2019).
98. Kelman, I., Ayeb-Karlsson, S., Schipper, L., Pelling, M. & Beck, S. *Learning from the History of Disaster Vulnerability and Resilience Research and Practice for Climate Change* (UK Alliance for Disaster Research, 2022).
99. Tol, R. S. The IPCC and the challenge of ex post policy evaluation. Preprint at https://arxiv.org/abs/2207.14724 (2022).

Acknowledgements

A.J.S. acknowledges the support of this work by the UK Natural Environment Research Council (Panorama DTP). J.C.M. acknowledges funding by the European Research Council under the European Union's Horizon 2020 Research and Innovation programme (GENIE Project, grant 951542).

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/s41558-023-01890-3.

Correspondence should be addressed to Anne J. Sietsma.

Peer review information *Nature Climate Change* thanks Vilhelm Verendel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.