



UNIVERSITY OF LEEDS

This is a repository copy of *The next generation of machine learning for tracking adaptation texts*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/208463/>

Version: Accepted Version

Article:

Sietsma, A.J., Ford, J.D. orcid.org/0000-0002-2066-3456 and Minx, J.C. (2023) The next generation of machine learning for tracking adaptation texts. *Nature Climate Change*, 14. pp. 31-39. ISSN 1758-678X

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

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

The Next Generation of Machine Learning for Tracking Adaptation Texts

Author Accepted Manuscript

This version of the article has been accepted for publication, after peer review and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: doi.org/10.1038/s41558-023-01890-3

Abstract

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 only deliver 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 by organisations like the IPCC could help lead the way.

Anne Sietsma^{1,2*}, James D. Ford¹, Jan C. Minx^{1,3}

¹ Priestley International Centre for Climate, University of Leeds, Leeds, United Kingdom

² Wageningen University, Wageningen, The Netherlands

³ Mercator Research Institute on Global Commons and Climate Change, Berlin, Germany

* Author to whom any correspondence should be addressed.

Email: anne.sietsma@wur.nl

Introduction

As the climate crisis continues, the need to respond to unavoidable impacts is increasing. This is known as adaptation – “[t]he process of adjustment to actual or expected climate and its effects in order to moderate harm or exploit beneficial opportunities”⁶, Box TS.1. Consequently, the importance of tracking adaptation also increases: good high-level overviews of adaptation efforts can help communities learn from each other and direct resources to where they are most needed⁷. Yet despite political urgency⁸, par. 74-77 and a multitude of frameworks and methodologies to do this⁹⁻¹², global tracking of the effectiveness and progress of adaptation actions has proven difficult^{13,14}. Current efforts tend to use proxies, including policies texts or funding flows^{9,10,15} or they document the state of scientific evidence, for example in the form of systematic maps^{4,5,16}. Such approaches do not measure adaptation outcomes – i.e. how much risk is being reduced – but they do provide insight into adaptation *processes* – i.e. where and how adaptation is taking place.

In the age of ‘Big Literature’¹⁹, part of the problem is the sheer volume and variety of the evidence base – e.g. tens of thousands of climate change papers are published each year²³, with adaptation evidence often stemming from case studies²⁴. Evidence of ongoing adaptation processes at the same time has become more widely available due to digitalisation and the increasing interest in adaptation in general⁵.

Additionally, there is an ongoing debate on how to define adaptation in general^{25,26}, and adaptation success^{9,13} 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 at the simultaneously capable of rapidly handling large and varied datasets, while still remaining sensitive to fine-grained distinctions and context-dependant meanings.

In theory, this is exactly what Machine Learning (ML) promises: taking human-like decision at scale rapidly^{27,28}.

Recently, that promise is increasingly put to the test in a first generation of articles that use ML methods to assess adaptation evidence in practice ^{5,7,29-31}. 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 former paints an overwhelmingly positive image of both current and future ML, the newly emerging experiences of those who have done 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 these personal experiences and related literature on tracking evidence through textual data; readers interested in ML applications for areas like image processing, remote sensing and risk modelling may wish consider additional literature also ³²⁻³⁴.

Overall, despite all concerns and criticisms voiced here, we continue to believe that ML could transform climate change research in general and adaptation research in particular. However, realising 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: 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 ^{27-29,35}. In the meantime, a first generation of application studies has emerged. A rapid review of the literature which either uses or substantially discusses the use of machine learning for adaptation evidence finds 54 relevant papers in Web of Science and Scopus (see Supplementary Materials for the protocol and included papers). Note that this excludes virtually all modelling and remote sensing work: although ML applications are gaining ground here too, ³²⁻³⁴ 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. Additionally, a substantial number of additional papers discusses 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.

Table 1: examples of studies using machine learning in an adaptation context. The full list of articles can be found in the Supplementary Materials. LDA = Latent Dirichlet Allocation; COP = Conference of the Parties, the main United Nations forum for climate change.

Reference	Dataset	ML method	Sample findings
Berrang-Ford et al., 2021 ⁷	Primary research articles indexed in Web of Science, Scopus or Medline	Supervised learning to select and categorise adaptation projects; pre-trained algorithm to extract geographic locations	Number of adaptation projects growing quickly, but research gaps persist: evidence is largely local and fragmented; evidence on transformational adaptation is limited.
Biesbroek et al., 2022	UNFCCC National Communications	Structural Topic Model (unsupervised learning)	Emphasis is on climate impacts, but shifting towards adaptation, governance and vulnerability; significant North-South differences
Boussalis et al., 2019 ³⁶	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)	Salience of adaptation is increasing overall; cities that are especially vulnerable discuss adaptation more, including some cities with Republican mayors.
Lesnikowski et al., 2019 ²⁹	Speeches at COPs; council minutes from 25 municipalities in Canada	LDA (unsupervised learning)	Disconnect in the discourse: Global South focusses on adaptation planning and feasibility; North on finance and overlaps with mitigation. Municipalities focus on extreme events and the built environment.
Zander et al., 2022 ³⁷	Hand-selected primary research articles on human mobility and the environment in Scopus	LDA (unsupervised learning) with a clustering algorithm (not machine learning)	Literature is diverse; Adaptation and impact literature relatively separate; focus on sudden hazards over long-term climate change.

All included papers are published in or after 2015 and 40 of these contain primary research; the remainder describes 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 uses a fairly small subset of methods, especially topic models and other clustering algorithms^{16,31,36-43}. *Topic models* are used to create an overview of a collection of texts by identifying and quantifying topics – i.e. 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 cited⁴⁴.

Such *unsupervised machine learning* seeks patterns in the input data without needing any kind of hand-labelled data. This means 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^{43,45}. Generally though, adaptation researchers opt for existing datasets, such as self-reported data from cities^{38,39}, or for data that is relatively easy to obtain in a structured manner, such as UNFCCC documents^{29,31,42} and especially scientific literature^{5,16,37,46}. Note that in all these cases, text-based data are used as a proxy for adaptation processes; 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^{5,16,31}.

Supervised machine learning by contrast is less commonly used for adaptation. These types of methods “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^{5,7,16,30}. 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 pre-determined task. The disadvantage however is equally clear: labelled data is rare and producing the required labels can be costly.

Still, some adaptation-relevant papers have used supervised methods ^{5,7,16,30,47,48}, and here again we find that these projects tend to rely on relatively well-established implementations, including Support Vector Machines ^{5,7,36,49} and Neural Nets ^{30,47,48}. The specific workings of these widely used models are discussed elsewhere ^{50,51} as both have a wide variety of use cases, but in the extant NLP-adaptation literature, they are typically used for categorising texts in a supervised manner. There are not many examples of *Large Language Models* (LLMs) or *Transformer Models* used for adaptation^{4,52,53} – the relatively small Transformers-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 “transfer learning” ⁵⁴. There are also *pre-trained models* ^{5,16,38,42,55} – i.e. ML algorithms that have been trained already on a different dataset for a specific task. Adaptation scholars generally use these for the relatively well-known tasks of sentiment analysis (i.e. 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 so far mostly provide 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 find 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.

Two next 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; 2) to reflect on both the strategic priorities field and the opportunities afforded by machine learning methods that are still emerging. We will discuss each in turn.

Promise meets practice

Subsequently, 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.

Table 2: Comparison of the theoretical promises of Machine Learning (ML) against the practical findings of projects which use ML to analyse adaptation evidence from text data, with some suggestions on how to move forward. Each of the rows is elaborated upon in the subsequent sections.

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.	Prioritise 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 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.

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 which use NLP methods to prioritise documents that are likely to be relevant^{56,57}. This can cut the review time in half or less^{56,58}. 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 sub-topics within adaptation ⁵⁹⁻⁶¹.

For even larger subjects, it may be better to train a new ML-model ⁵⁶. 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 2-4 thousand screened articles. This implies a significant amount of time labelling articles – even if one article would cost one minute to label, that is around 50 hours – but if the complete search returned, for example, 15 thousand documents, manual screening of all abstracts would take roughly 5 times longer still. Larger searches may benefit even more ^{4,5} though care should be taken that 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 the time investment is broadly similar to for example bibliometric analyses; however those rely on relatively crude heuristics (e.g. keywords or the number of times a single word is used), whereas methods like 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 ⁶²– 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 of what an appropriate size dataset is. 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. Additionally, 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, machine learning tools therefore are useful if there are at least a few hundred documents on the subject of interest.^{4,5,16} Note however that this threshold may be lowered in the near future; LLMs in particular are getting better at a wide range of tasks, which include zero-shot learning⁶³ and generating synthetic data⁶⁴. More generally, their “emergent abilities”⁶⁵ mean they are capable of performing tasks which 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 like ChatGTP or Bard, or when training Transformer models like 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 (API) 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, adaptation tracking literature to date tends to focus on well-established data sources. A common suggestion is that future research should include more diverse sources, especially 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 Initiative⁷ relied on a large team of 126 researchers, but even this proved insufficient

to systematically include grey literature. Relatedly, Hsu and Rauber ³⁹ provide one of the few examples where a substantial number of databases are combined, but they still caution that their data largely originates from Europe and “is limited by the lack of time-series data, regular and repeated reporting on climate actions, strategies, and policies” (p. 9). 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 standardised databases ⁶⁶ and methods. Researchers could, for example, use web scrapers to specifically target government websites of areas where traditional data coverage is poor (e.g. many areas in the Global South). Combining different sources will require additional experimentation, for example with automated summarisers to create document sets of a more homogenous length, by translating non-English data automatically, or by using multi-language models. For adaptation tracking in scientific texts, we see a large role for database providers and libraries ⁶⁷ who could more systematically index non-academic sources and make them available in a standardised 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 machine learning 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, 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, this is not only untrue but can also 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 ^{27,29}. 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 – though it may be some time yet before this is sufficiently reliable ⁶⁸. 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 – i.e. the percentage of correct classifications – but 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. It is typically around 70-90% for binary problems ^{4,5,23,30}, but it may be much lower for complex tasks ^{52,69}. Unless the reviewer has a background in ML, they will likely 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 ^{47,70} or place accuracy numbers in the supplementary materials e.g. ^{7,53,71} 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 neither be able to ask the right questions, nor to operationalise and execute the projects adequately. To give an NLP example, consider how one might find documents on adaptation. Given that a large dataset is less of a concern for NLP methods, one might opt for a simple query with general terms, combining different forms of “climate” and “adaptation” (e.g. *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 previous 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 (e.g. *mitigat**) in the search as “risk mitigation” is sometimes used instead of “adaptation”^{59,72}. 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 interpret²⁹. There are two caveats here: first, some quantitative measures for topic model quality do exist⁷³; second, some use topic models purely to explore the 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 contextualised and critically analysed, which requires domain-specific expertise.

Collaborations between computer scientists and domain experts may help bring the required knowledge together. For academia, the *climatechange.ai* community³⁵ has set up climate change tracks at computer science conferences. Conversely, we would urge the organisers of adaptation conferences to also actively invite the machine learning community (e.g. Adaptation Futures or European Climate Change Adaptation). Universities and individual academics can foster interdisciplinary collaborations too; for reasons of space, we will point to recent overview by Lyall⁷⁴ 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. Interdisciplinarity research broadly recognises that combining different domain-specific epistemologies is often difficult and time consuming.^{75,76} In ML projects specifically, the model parameters will influence the outcome and are dependent on the data. This means 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 artifacts 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 help create this in theory; however, training materials on ML, including NLP, have been widely available for quite some time, yet adaptation applications are few and far between. We therefore believe adaptation organisations should focus first on improving “machine learning literacy” – i.e. 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 helps adaptation practitioners recognise opportunities for ML in their own work, while also ensuring that results can be fruitfully discussed and that researches 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 ⁷⁷ should be the starting point of such guidelines, both to increase transparency and to accelerate progress. In sum, focussing on the basis can help create space for a new generation of researchers to develop shared goals and norms.

Models repeat biases and conventional wisdoms

The third promise 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, rather than assume them” ⁷⁸ p. 1066. This quote and similar formulations are used to make two closely related points: first, it suggests topic models are less biased e.g. ^{29,37,39}; second, these tools would lead to new insights as they can “uncover hidden patterns” ⁷⁹, p. 136 and “identify facts and relationships that would otherwise remain buried” ⁵⁵, p. 4. 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 shift where bias is introduced ⁸⁰. Running a topic model on biased data will still have 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^{7,16 52}, 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 – e.g. there is an ongoing and politically-charged debate on what differentiates adaptation from general development²⁵, 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 pre-trained models. There are a variety of methods to quantify or adjust for biases in these models,^{81,82} 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 resources⁸³. Meanwhile, LLM’s bias around gender, race, and religion, among others, are well documented^{84,85}. 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. Since 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.

Table 3 (next page): *examples of potential bias in a large language model. Examples were generated using a transformer-based model. Such models are created by training on large sets of documents – here we use climateBERT, of 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 (i.e. “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.*

Prompt	Most likely	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 ^{86,87}
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 who need to work together to overcome climate risks. Mitigation is something one 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, while 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 2.5% probability in the first example; “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, likely because a disproportionate amount of climate research originates here ⁵ (September scores a 6.7% probability in the second example).
The summer starts in [blank].	June (13.9%) May (13.1%)	

Similarly, ML methods can certainly be used to generate novel insights, but it is “data hubris” ⁸⁸ to think that with sufficient data and algorithms, such insights will simply reveal themselves. Even the most cutting-edge NLP models have been called “stochastic parrots” ⁸⁹ which lack critical thinking and cannot distinguish between what is widely *stated* versus what is widely (dis)*proven*. Further, in the rare cases where ML outcomes are compared to expert opinions (e.g. ^{5,38}), 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 – e.g. comparing twitter discourse to academic publishing ⁹⁰, overlaying grid-cell climate models and observations with the location and topics of climate impact studies ⁴. As noted earlier, however, heterogenous datasets will take considerable effort to create.

Secondly, 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 only result in well-known broad trends. Examples include looking for shifts in national reporting post-Paris Agreement and finding such shifts are barely perceptible ⁹¹; and investigating whether the intended interlinkages between different Sustainable Development Goals are also found in practice ⁹². In this light, it is worth noting that formal hypothesis testing and error ranges are often not reported in ML-assisted syntheses (by contrast: ^{4,5}).

Treating Machine Learning as a Transformation

So far, we have focussed 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 – i.e. 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 co-exist. But they will compete for resources. It is therefore worth thinking critically on the types of insights 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 been done already and play to the strengths of the method.

An underrated element to 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 updating or fine-tuning the model may be necessary). This makes ML especially useful for the types of repetitive tasks that form the foundation of many adaptation projects, such as finding adaptation-relevant passages in policy documents ³⁰, linking adaptation evidence to locations and impact models ⁴, identifying knowledge gaps ^{5,93}, or, more ambitiously, creating a standardised global database of adaptation interventions⁶⁶. 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 alleviate the inequal access to computational power.

In addition, if one re-runs the whole “pipeline” of an NLP-based tracking system at regular intervals, one could create a near-real-time overview of the evidence. To some, self-updating tracking systems may seem futuristic, but the reality is that technologically, this is feasible already. So-called “living” evidence approaches have recently gained popularity, especially in the health sciences ^{94,95}. Although living reviews may still have a manual component (e.g. 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 ⁹⁶.

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 likely 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, time-period – which can then form the basis for an easily searchable platform and interactive graphics. There are recent examples of projects which incorporate some of these suggestions, such as Climate Policy Radar, ⁹⁷ but overall, interactive platforms are rare. Scientific outputs especially often take the form of a

Textbox 1: why organisations like the IPCC should create flagship projects

A pressing example of the need for novel tracking methods is the IPCC, which is entering its 7th assessment period. Their mandate to synthesise all available climate research is increasingly difficult to meet: the latest Working Group II report relies on hundreds of authors and includes over 34 thousand 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 ²⁻⁴; similarly, because the reports are only issued every few years, they can lag behind the research frontier on key emerging issues ⁵.

Building on much older critiques ^{17,18}, 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 ^{20,21} (e.g. on policy implementation ²²). In our view, it is clear that the IPCC at least needs to innovate, and the organisation 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 policy maker 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.

At the community level, such a project could have positive knock-on effects, as it would help establish a “gold standard” for ML work in climate change, showcase what is currently possible and attract additional talent, which, over time, will pay dividends.

There are other organisations too, which could play such a role, for example UNEP, which writes the Adaptation Gap Reports. Depending on political circumstances, this may be more expedient, though the IPCC is a trusted source, which would help 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.

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. Since datasets are relegated to the supplementary materials, 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 upon a “gold standard” interactive and living evidence platform in Textbox 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^{98,99}. Researchers need to tread a careful line here: while LLMs can remove common bottle necks, in the absence of robust domain-specific evaluation metrics, human-in-the-loop systems may be a more prudent path.¹⁰⁰ Even so, making the most of the full breadth of ML tools will require some foresight and planning, especially around identifying the types of tasks 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 to adapt effectively. But this is no small task; the adaptation community has work to do.

Acknowledgements

AJS acknowledges that this work was supported by the UK Natural Environment Research Council (Panorama DTP). JCM 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.

Data availability

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

Supplementary materials

The supplementary materials contain 1) the protocol for our rapid review, and 2) the resulting table with the details of all documents considered, including reasons for exclusion. For relevant documents, the overall sentiment towards machine learning is given, alongside the data and machine learning methods used.

References

- 1 IPCC. 8 (2021).
- 2 Minx, J. C., Callaghan, M., Lamb, W. F., Garard, J. & Edenhofer, O. Learning about climate change solutions in the IPCC and beyond. *Environmental Science & Policy* 77, 252-259 (2017). <https://doi.org/10.1016/j.envsci.2017.05.014>
- 3 Berrang-Ford, L. *et al.* Editorial: Evidence synthesis for accelerated learning on climate solutions. *Campbell Systematic Reviews* 16 (2020). <https://doi.org/10.1002/cl2.1128>
- 4 Callaghan, M. *et al.* Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nature Climate Change* 11, 966-972 (2021). <https://doi.org/10.1038/s41558-021-01168-6>
- 5 Sietsma, A. J., Ford, J. D., Callaghan, M. W. & Minx, J. C. Progress in climate change adaptation research. *Environmental Research Letters* 16, 054038 (2021).
- 6 IPCC. *Climate Change 2022: Impacts, Adaptation and Vulnerability*. (Cambridge University Press, 2022).
- 7 Berrang-Ford, L. *et al.* A systematic global stocktake of evidence on human adaptation to climate change. *Nature Climate Change* 11, 989-1000 (2021). <https://doi.org/10.1038/s41558-021-01170-y>
- 8 UNFCCC. in *FCCC/PA/CMA/2022/L.21* (ed Conference of the Parties serving as the meeting of the Parties to the Paris Agreement) (United Nations Framework Convention on Climate Change, Sharm el-Sheikh, 2022).
- 9 Berrang-Ford, L. *et al.* Tracking global climate change adaptation among governments. *Nature Climate Change* 9, 440-449 (2019). <https://doi.org/10.1038/s41558-019-0490-0>
- 10 Craft, B. & Fisher, S. Measuring the adaptation goal in the global stocktake of the Paris Agreement. *Climate Policy* 18, 1203-1209 (2018). <https://doi.org/10.1080/14693062.2018.1485546>

- 11 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. *Climate Risk Management*
12 **35**, 100401 (2022). [https://doi.org:https://doi.org/10.1016/j.crm.2022.100401](https://doi.org/10.1016/j.crm.2022.100401)
- 13 Olhoff, A., Väänänen, E. & Dickson, B. in *Resilience* (eds Zinta Zommers & Keith
Alverson) 51-61 (Elsevier, 2018).
- 14 Dilling, L. *et al.* Is adaptation success a flawed concept? *Nature Climate Change* **9**,
572-574 (2019). [https://doi.org:10.1038/s41558-019-0539-0](https://doi.org/10.1038/s41558-019-0539-0)
- 15 Leiter, T. & Pringle, P. in *Adaptation metrics: Perspectives on measuring,
aggregating and comparing adaptation results* (eds Lars Christiansen, Gerardo
Martinez, & Prakriti Naswa) 29-48 (UNEP DTU, Copenhagen, 2018).
- 16 UNEP. Adaptation Gap Report 2022: Too Little, Too Slow – Climate adaptation failure
puts world at risk., (United Nations Environment Programme, Nairobi, 2022).
- 17 Berrang-Ford, L. *et al.* Systematic mapping of global research on climate and health: a
machine learning review. *The Lancet Planetary Health* **5**, e514-e525 (2021).
[https://doi.org:https://doi.org/10.1016/S2542-5196\(21\)00179-0](https://doi.org/10.1016/S2542-5196(21)00179-0)
- 18 Peticrew, M. & McCartney, G. Using Systematic Reviews to Separate Scientific from
Policy Debate Relevant to Climate Change. *American Journal of Preventive Medicine*
19 **40**, 576-578 (2011). [https://doi.org:https://doi.org/10.1016/j.amepre.2010.12.022](https://doi.org/10.1016/j.amepre.2010.12.022)
- 20 Tol, R. S. J. Regulating knowledge monopolies: the case of the IPCC. *Climatic Change*
108, 827 (2011). [https://doi.org:10.1007/s10584-011-0214-6](https://doi.org/10.1007/s10584-011-0214-6)
- 21 Nunez-Mir, G. C., Iannone Iii, B. V., Pijanowski, B. C., Kong, N. & Fei, S. Automated
content analysis: addressing the big literature challenge in ecology and evolution.
Methods in Ecology and Evolution **7**, 1262-1272 (2016).
[https://doi.org:https://doi.org/10.1111/2041-210X.12602](https://doi.org/10.1111/2041-210X.12602)
- 22 Provost, G. Rigorous and relevant: Applying lessons from the history of IPCC special
reports to the Post-Paris agreement world. *Harv. Envtl. L. Rev.* **43**, 507 (2019).
- 23 Kelman, I., Ayeb-Karlsson, S., Schipper, L., Pelling, M. & Beck, S. (UK Alliance for
Disaster Research, 20 May 2022, 2022).
- 24 Tol, R. S. The IPCC and the challenge of ex post policy evaluation. *arXiv preprint*
arXiv:2207.14724 (2022).
- 25 Callaghan, M. W., Minx, J. C. & Forster, P. M. A topography of climate change research.
Nature Climate Change **10**, 118-123 (2020). [https://doi.org:10.1038/s41558-019-0684-5](https://doi.org/10.1038/s41558-019-0684-5)
- 26 Nalau, J. & Verrall, B. Mapping the evolution and current trends in climate change
adaptation science. *Climate Risk Management* **32** (2021).
[https://doi.org:10.1016/j.crm.2021.100290](https://doi.org/10.1016/j.crm.2021.100290)
- 27 Schipper, E. L. F., Tanner, T., Dube, O. P., Adams, K. M. & Huq, S. The debate: Is global
development adapting to climate change? *World Development Perspectives* **18**,
100205 (2020). [https://doi.org:https://doi.org/10.1016/j.wdp.2020.100205](https://doi.org/10.1016/j.wdp.2020.100205)
- 28 Siders, A. R. Adaptive capacity to climate change: A synthesis of concepts, methods,
and findings in a fragmented field. *WIREs Climate Change* **10**, e573 (2019).
[https://doi.org:10.1002/wcc.573](https://doi.org/10.1002/wcc.573)
- 29 Ford, J. D. *et al.* Big data has big potential for applications to climate change
adaptation. *Proceedings of the National Academy of Sciences* **113**, 10729-10732
(2016).
- 30 Cheong, S. M., Sankaran, K. & Bastani, H. Artificial intelligence for climate change
adaptation. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge
Discovery*, e1459 (2022).
- 31 Lesnikowski, A. *et al.* Frontiers in data analytics for adaptation research: Topic
modeling. *Wiley Interdisciplinary Reviews: Climate Change* **10**, e576 (2019).
- 32 Biesbroek, R., Badloe, S. & Athanasiadis, I. N. Machine learning for research on climate
change adaptation policy integration: an exploratory UK case study. *Regional
Environmental Change* **20**, 85 (2020). [https://doi.org:10.1007/s10113-020-01677-8](https://doi.org/10.1007/s10113-020-01677-8)

- 31 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). *Climate Policy* **22**, 97-111 (2022). <https://doi.org:10.1080/14693062.2021.2018986>
- 32 Zennaro, F. *et al.* Exploring machine learning potential for climate change risk assessment. *Earth-Science Reviews* **220**, 103752 (2021). <https://doi.org:https://doi.org/10.1016/j.earscirev.2021.103752>
- 33 Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A. & Kumar, V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering* **31**, 1544-1554 (2018).
- 34 Munawar, H. S., Hammad, A. W. A. & Waller, S. T. A review on flood management technologies related to image processing and machine learning. *Automation in Construction* **132**, 103916 (2021). <https://doi.org:https://doi.org/10.1016/j.autcon.2021.103916>
- 35 Rolnick, D. *et al.* Tackling Climate Change with Machine Learning. *ACM Comput. Surv.* **55**, Article 42 (2022). <https://doi.org:10.1145/3485128>
- 36 Boussalis, C., Coan, T. G. & Holman, M. R. Communicating Climate Mitigation and Adaptation Efforts in American Cities. *Climate* **7**, 45 (2019).
- 37 Zander, K. K. *et al.* Topic modelling exposes disciplinary divergence in research on the nexus between human mobility and the environment. *Humanities and Social Sciences Communications* **9**, 34 (2022). <https://doi.org:10.1057/s41599-022-01038-2>
- 38 Fu, X., Li, C. & Zhai, W. Using Natural Language Processing to Read Plans. *Journal of the American Planning Association*, 1-12 (2022). <https://doi.org:10.1080/01944363.2022.2038659>
- 39 Hsu, A. & Rauber, R. Diverse climate actors show limited coordination in a large-scale text analysis of strategy documents. *Communications Earth & Environment* **2**, 30 (2021). <https://doi.org:10.1038/s43247-021-00098-7>
- 40 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. *Applied Sciences* **9**, 3859 (2019).
- 41 Paulvannan Kanmani, A., Obringer, R., Rachunok, B. & Nateghi, R. Assessing Global Environmental Sustainability Via an Unsupervised Clustering Framework. *Sustainability* **12**, 563 (2020).
- 42 Valero, S. D., Emandi, R., Encarnacion, J., Kaul, S. & Seck, P. Utilizing big data to measure key connections between gender and climate change 1. *Statistical Journal of the IAOS* **38**, 973-994 (2022). <https://doi.org:10.3233/SJI-220964>
- 43 Lynam, T. Exploring social representations of adapting to climate change using topic modeling and Bayesian networks. *Ecology and Society* **21** (2016). <https://doi.org:10.5751/ES-08778-210416>
- 44 Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent dirichlet allocation. *Journal of machine Learning research* **3**, 993-1022 (2003).
- 45 Tvinnerheim, 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. *Global Environmental Change* **46**, 34-41 (2017). <https://doi.org:https://doi.org/10.1016/j.gloenvcha.2017.06.005>
- 46 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).
- 47 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 Climate* **45**, 101266 (2022). <https://doi.org:https://doi.org/10.1016/j.uclim.2022.101266>
- 48 Canon, M. J., Satuito, A. & Sy, C. in *2018 International Symposium on Computer, Consumer and Control (IS3C)*. 237-241.

- 49 Salam, R. *et al.* Nexus between vulnerability and adaptive capacity of drought-prone rural households in northern Bangladesh. *Natural Hazards* **106**, 509-527 (2021). <https://doi.org:10.1007/s11069-020-04473-z>
- 50 Abiodun, O. I. *et al.* State-of-the-art in artificial neural network applications: A survey. *Heliyon* **4**, e00938 (2018). <https://doi.org:https://doi.org/10.1016/j.heliyon.2018.e00938>
- 51 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). <https://doi.org:https://doi.org/10.1016/j.neucom.2019.10.118>
- 52 Sietsma, A. J. *et al.* Climate Change Adaptation Policy Across Scales: A Machine Learning Evidence Map. *One Earth* (Under review).
- 53 Binger, J. A., Kraus, M., Leippold, M. & Webersinke, N. Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures. *Finance Research Letters* **47**, 102776 (2022). <https://doi.org:https://doi.org/10.1016/j.frl.2022.102776>
- 54 Gillioz, A., Casas, J., Mugellini, E. & Khaled, O. A. in *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*. 179-183.
- 55 Huo, F. *et al.* Using big data analytics to synthesize research domains and identify emerging fields in urban climatology. *WIREs Climate Change* **12**, e688 (2021). <https://doi.org:https://doi.org/10.1002/wcc.688>
- 56 Khalil, H., Ameen, D. & Zarnegar, A. Tools to support the automation of systematic reviews: a scoping review. *Journal of Clinical Epidemiology* **144**, 22-42 (2022). <https://doi.org:10.1016/j.jclinepi.2021.12.005>
- 57 Marshall, C., Sutton, A., O'Keefe, H. & Johnson, E. (2022).
- 58 Gates, A. *et al.* Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. *Systematic Reviews* **8**, 278 (2019). <https://doi.org:10.1186/s13643-019-1222-2>
- 59 Bisaro, A., Roggero, M. & Villamayor-Tomas, S. Institutional Analysis in Climate Change Adaptation Research: A Systematic Literature Review. *Ecological Economics* **151**, 34-43 (2018). <https://doi.org:https://doi.org/10.1016/j.ecolecon.2018.04.016>
- 60 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. *Environmental Research Letters* **16**, 073001 (2021).
- 61 Naulleau, A., Gary, C., Prévot, L. & Hossard, L. Evaluating strategies for adaptation to climate change in grapevine production—A systematic review. *Frontiers in plant science* **11**, 607859 (2021).
- 62 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 Research & Social Science* **70**, 101691 (2020). <https://doi.org:https://doi.org/10.1016/j.erss.2020.101691>
- 63 Wu, P. Y., Tucker, J. A., Nagler, J. & Messing, S. Large language models can be used to estimate the ideologies of politicians in a zero-shot learning setting. *arXiv preprint arXiv:2303.12057* (2023).
- 64 Hämäläinen, P., Tavast, M. & Kunnari, A. in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-19.
- 65 Wei, J. *et al.* Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682* (2022).
- 66 Canales, N., Klein, R. J. T., Bakhtaoui, I. & Macura, B. Assessing adaptation progress for the global stocktake. *Nature Climate Change* (2023). <https://doi.org:10.1038/s41558-023-01656-x>
- 67 Marsolek, W., Farrell, S. L., Kelly, J. A. & Cooper, K. Grey literature: Advocating for diverse voices, increased use, improved access, and preservation. *College & Research Libraries News*; Vol 82, No 2 (2021): February (2021). <https://doi.org:10.5860/crln.82.2.58>

- 68 Poldrack, R. A., Lu, T. & Beguš, G. AI-assisted coding: Experiments with GPT-4. *arXiv preprint arXiv:2304.13187* (2023).
- 69 Corringham, T. *et al.* in *ICML 2021 Workshop on Tackling Climate Change with Machine Learning* (2021).
- 70 Manandhar, A. *et al.* Machine Learning to Evaluate Impacts of Flood Protection in Bangladesh, 1983–2014. *Water* **12**, 483 (2020).
- 71 Sachdeva, S., Hsu, A., French, I. & Lim, E. A computational approach to analyzing climate strategies of cities pledging net zero. *npj Urban Sustainability* **2**, 21 (2022). <https://doi.org:10.1038/s42949-022-00065-x>
- 72 Kim, B. J., Jeong, S. & Chung, J.-B. Research trends in vulnerability studies from 2000 to 2019: Findings from a bibliometric analysis. *International Journal of Disaster Risk Reduction* **56**, 102141 (2021). <https://doi.org:https://doi.org/10.1016/j.ijdrr.2021.102141>
- 73 Jacobs, T. & Tschötschel, R. Topic models meet discourse analysis: a quantitative tool for a qualitative approach. *International Journal of Social Research Methodology* **22**, 469-485 (2019). <https://doi.org:10.1080/13645579.2019.1576317>
- 74 Lyall, C. *Being an interdisciplinary academic: How institutions shape university careers.* (Springer Nature, 2019).
- 75 MacLeod, M. What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. *Synthese* **195**, 697-720 (2018). <https://doi.org:10.1007/s11229-016-1236-4>
- 76 Miller, T. R. *et al.* Epistemological Pluralism
Reorganizing Interdisciplinary Research. *Ecology and Society* **13** (2008).
- 77 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. *Environmental Evidence* **7**, 7 (2018). <https://doi.org:10.1186/s13750-018-0121-7>
- 78 Roberts, M. E. *et al.* Structural topic models for open-ended survey responses. *American journal of political science* **58**, 1064-1082 (2014).
- 79 Miglionico, A. The Use of Technology in Corporate Management and Reporting of Climate-Related Risks. *European Business Organization Law Review* **23**, 125-141 (2022). <https://doi.org:10.1007/s40804-021-00233-z>
- 80 Hovy, D. & Prabhunoye, S. Five sources of bias in natural language processing. *Language and Linguistics Compass* **15**, e12432 (2021). <https://doi.org:https://doi.org/10.1111/lnc3.12432>
- 81 Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183-186 (2017). <https://doi.org:10.1126/science.aal4230>
- 82 Guo, Y., Yang, Y. & Abbasi, A. 1012-1023 (Association for Computational Linguistics).
- 83 Xu, F. F., Alon, U., Neubig, G. & Hellendoorn, V. J. in *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming.* 1-10.
- 84 Garrido-Muñoz, I., Montejo-Ráez, A., Martínez-Santiago, F. & Ureña-López, L. A. A Survey on Bias in Deep NLP. *Applied Sciences* **11**, 3184 (2021).
- 85 Magee, L., Ghahremanlou, L., Soldatic, K. & Robertson, S. Intersectional bias in causal language models. *arXiv preprint arXiv:2107.07691* (2021).
- 86 Huyer, S. & Gumucio, T. Going Back to the Well: Women, Agency, and Climate Adaptation. *World Journal of Agriculture and Soil Science* **5**, 1-3 (2020).
- 87 Wester, M. & Lama, P. D. in *Climate hazards, disasters, and gender ramifications* 67-85 (Routledge, 2019).
- 88 Lazer, D., Kennedy, R., King, G. & Vespignani, A. The parable of Google Flu: traps in big data analysis. *Science* **343**, 1203-1205 (2014).
- 89 Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* 610-623.

- 90 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. *Journal of Informetrics* **13**, 695-707 (2019). <https://doi.org/10.1016/j.joi.2019.03.008>
- 91 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. *Regional Environmental Change* **23**, 1-14 (2023). <https://doi.org/10.1007/s10113-023-02113-3>
- 92 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. *Scientific Reports* **11**, 22427 (2021). <https://doi.org/10.1038/s41598-021-01801-6>
- 93 Berrang Ford, L. *et al.* Mapping climate-health evidence: Using machine-learning to map the links between climate change and health. 192 (Foreign, Commonwealth and Development Office, London, 2021).
- 94 Elliott, J. H. *et al.* Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of Clinical Epidemiology* **91**, 23-30 (2017). <https://doi.org/10.1016/j.jclinepi.2017.08.010>
- 95 Millard, T. *et al.* Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation. *Systematic Reviews* **8**, 325 (2019). <https://doi.org/10.1186/s13643-019-1248-5>
- 96 Khalil, H., Tamara, L., Rada, G. & Akl, E. A. Challenges of evidence synthesis during the 2020 COVID pandemic: a scoping review. *Journal of Clinical Epidemiology* **142**, 10-18 (2022). <https://doi.org/10.1016/j.jclinepi.2021.10.017>
- 97 Climate Policy Radar. *Climate Policy Radar App*, <<https://app.climatepolicyradar.org/>> (2023).
- 98 Liu, Y. *et al.* Summary of chatgpt/gpt-4 research and perspective towards the future of large language models. *arXiv preprint arXiv:2304.01852* (2023).
- 99 Floridi, L. & Chiriatti, M. GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines* **30**, 681-694 (2020). <https://doi.org/10.1007/s11023-020-09548-1>
- 100 Debnath, R., Creutzig, F., Sovacool, B. K. & Shuckburgh, E. Harnessing human and machine intelligence for planetary-level climate action. *npj Climate Action* **2**, 20 (2023). <https://doi.org/10.1038/s44168-023-00056-3>