Environmental Data Science (2022), 1: e6, 1–15
doi:10.1017/eds.2022.5

POSITION PAPER

Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science

Amy McGovern[1,6,*], Imme Ebert-Uphoff[2,3,6], David John Gagne II[4,6] and Ann Bostrom[5,6]

[1]School of Computer Science and School of Meteorology, University of Oklahoma, Norman, OK 73019, USA
[2]Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA
[3]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80523, USA
[4]Computational and Information Systems Laboratory, National Center for Atmospheric Research, Boulder, CO 80301, USA
[5]Evans School of Public Policy & Governance, University of Washington, Seattle, WA 98195, USA
[6]NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography
*Corresponding author. E-mail: amcgovern@ou.edu

Received: 14 December 2021; **Revised:** 22 February 2022; **Accepted:** 23 February 2022

Keywords: Artificial intelligence; climate; ethics; weather

Abstract

Given the growing use of Artificial intelligence (AI) and machine learning (ML) methods across all aspects of environmental sciences, it is imperative that we initiate a discussion about the ethical and responsible use of AI. In fact, much can be learned from other domains where AI was introduced, often with the best of intentions, yet often led to unintended societal consequences, such as hard coding racial bias in the criminal justice system or increasing economic inequality through the financial system. A common misconception is that the environmental sciences are immune to such unintended consequences when AI is being used, as most data come from observations, and AI algorithms are based on mathematical formulas, which are often seen as objective. In this article, we argue the opposite can be the case. Using specific examples, we demonstrate many ways in which the use of AI can introduce similar consequences in the environmental sciences. This article will stimulate discussion and research efforts in this direction. As a community, we should avoid repeating any foreseeable mistakes made in other domains through the introduction of AI. In fact, with proper precautions, AI can be a great tool to help *reduce* climate and environmental injustice. We primarily focus on weather and climate examples but the conclusions apply broadly across the environmental sciences.

Impact Statement

This position paper discusses the need for the environmental sciences community to ensure that they are developing and using artificial intelligence (AI) methods in an ethical and responsible manner. This paper is written at a general level, meant for the broad environmental sciences and earth sciences community, as the use of AI methods continues to grow rapidly within this community.

1. Motivation

Artificial intelligence (AI) and machine learning (ML) have recently exploded in popularity for a wide variety of environmental science applications (e.g., McGovern et al., 2019; Reichstein et al., 2019;

Gagne et al., 2020; Gensini et al., 2021; Hill and Schumacher, 2021; Lagerquist et al., 2021; Schumacher et al., 2021). Like other fields, environmental scientists are seeking to use AI/ML to build a linkage from raw data, such as satellite imagery and climate models, to actionable decisions.

While the increase in applications of AI can bring improved predictions, for example, for a variety of high-impact events, it is also possible for AI to unintentionally do more harm than good if it is not developed and applied in an ethical and responsible manner. This has been demonstrated in a number of high-profile cases in the news outside of weather or climate (e.g., O'Neil, 2016; Benjamin, 2019; Kantayya, 2020). We argue that the potential for similar issues exists in environmental science applications and demonstrate how AI/ML methods could go wrong for these application areas.

On the other hand, AI can also be helpful for environmental sustainability. AI is already being applied to enable the automated monitoring of our ecosystems to support accountability for climate justice. Applications include monitoring land cover changes to detect deforestation (Karpatne et al., 2016; Mithal et al., 2018), counting populations of endangered species in very-high resolution satellite data (Duporge et al., 2021), and tracking bird populations in radar data (Chilson et al., 2019; Lin et al., 2019). Automated monitoring of retrospective and real-time datasets can help interested parties to monitor environmental trends and respond appropriately.

In our listing of how AI can go wrong for the environmental sciences, we specifically want to address environmental justice and injustice. According to the EPA (EPA, 2022), "environmental justice (EJ) is the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation and enforcement of environmental laws, regulations and policies." In the following sections, many aspects of the ways in which AI can go wrong (and the ways to address it) center around this definition.

There are both instrumental and consequential as well as principled ethical perspectives such as duties and virtues for AI environmental scientists to consider. These are entangled in the biases and pitfalls this paper explores. Debiasing has the potential to address both.

Note, we assume the reader is generally familiar with the concepts of AI and ML and do not focus on any specific AI/ML methods but instead on the applications across environmental sciences. For brevity, we refer to AI/ML as AI throughout the rest of the paper.

2. How AI can Go Wrong for Environmental Sciences

Although Arthur C. Clarke's famous line "Any sufficiently advanced technology is indistinguishable from magic" was not directed at AI, it may seem that way to many new users of AI. It seems as if one simply has to give an AI method data and use a premade package and an amazing predictive model results. The problem with this approach is that AI is not magic and can be led astray in a number of ways, which we outline here.

Box 1 provides a nonexhaustive list of ways in which AI can go wrong for environmental sciences, and other issues that can arise. We discuss this list in more depth below and provide examples from the environmental sciences to illustrate them.

2.1. Issues related to training data

AI models follow the well-known computer science adage: Garbage in, garbage out. If the training data are not representative of what we actually want to use the AI model for, then it will produce models that do not accomplish our goals. Because training data are key, issues related to them are particularly common and hard to avoid. First of all, as we demonstrate in the examples below, it is extremely difficult to create a dataset that does not have some kind of bias or other shortcoming, even if it was developed with great care and with the best intentions. Secondly, while the data developers might be aware of many (but usually not all) of the shortcomings, the scientist *using* the data to train AI models might be completely unaware of them, since there is no standard (yet!) in the community to document these shortcomings and include them with the data. Furthermore, whether a dataset is biased also depends on the application. It might be

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

Box 1. A nonexhaustive list of issues that can arise though the use of AI for environmental science applications.

Ways in which AI can go wrong for environmental sciences

Issues related to training data:

1. Non-representative training data, including lack of geo-diversity
2. Training labels are biased or faulty
3. Data is affected by adversaries

Issues related to AI models:

1. Model training choices
2. Algorithm learns faulty strategies
3. AI learns to fake something plausible
4. AI model used in inappropriate situations
5. Non-trustworthy AI model deployed
6. Lack of robustness in the AI model

Other issues related to workforce and society:

1. Globally applicable AI approaches may stymie burgeoning efforts in developing countries.
2. Lack of input or consent on data collection and model training
3. Scientists might feel disenfranchised.
4. Increase of CO_2 emissions due to computing

perfectly fine to use a dataset for one purpose, but not for another. Lastly, there is no established set of tests to check for the most common biases in environmental science datasets.

It is important to understand that when AI models are trained on biased data, they *inherit* those biases. This phenomenon is known as *coded bias* (O'Neil, 2016; Kantayya, 2020) and is easy to understand—if a hail dataset shows hail occurring only in populated areas (see Example 1(a)), the AI system trained on this dataset is likely to also predict hail only in highly populated areas, making the AI model just as biased to population density as the data it was trained on. Since established algorithms are often used in regions or under circumstances other than the origin of the data, such an algorithm can even perpetuate the bias beyond the regions and circumstances it was trained on.

2.1.1. Nonrepresentative training data

Statistical distribution of training, testing, and validation data: If the training data are nonrepresentative statistically, the output is likely to be biased. This could include geographic or population biases (discussed in more detail in Section 2.1.2) but it could also include unintentional temporal biases. For example, if an AI model is trained on data from 1900 to 1970, 1970 to 2000 is used as a validation set, and then 2000 to 2020 is used as a test set, the model may not be properly accounting for climate change. Sensor limitations provide other examples. Many sensors require sunlight for high quality observations, and thus many phenomena are under-observed at night. Similarly, inferring certain surface conditions from satellites using passive sensors tends to be easier for clear sky conditions.

Lack of geo-diversity: In order to enable AI algorithms to facilitate environmental justice we thus need to ensure that different populations are well represented in their training data. This includes ensuring that

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

the data are diverse geographically, including addressing population biases and biases in sensor placements that could affect the ability to obtain a statistically representative training set. For example, the national radar network has coverage gaps which can inadvertently under-represent some populations, as indicated in Figure 1 (Sillin, 2021; Shepherd, 2021). We should explicitly take population representation into account for the design of future sensor networks and seek to close gaps in existing sensor networks by placing additional sensors in strategic locations. Furthermore, we can put AI to good use by developing AI algorithms to estimate sensor values at the missing locations based on other sensor types, such as generating synthetic radar imagery from geostationary satellite imagery (Hilburn et al., 2021).

Lastly, because environmental sciences are characterized by a superposition of various processes at very different temporal and spatial scales, it is challenging to have a dataset fully representative of all potential processes to be analyzed and inferred by AI/ML methods; some processes may be neglected in the process.

2.1.2. Training labels are biased or faulty

Human generated datasets: Obtaining data for environmental science prediction tasks is often quite challenging, both in terms of finding the data and in terms of identifying a complete dataset. For example, if one was to predict hail or tornadoes but only to use human reports, there is a known bias toward areas with higher population (Allen and Tippett, 2015; Potvin et al., 2019) as shown in Figure 2. If more people live in an area, there is a higher chance that someone observes and reports a hail or tornado event. This might bias the AI model to over-predict urban hail/tornadoes and under-predict rural hail/tornadoes.

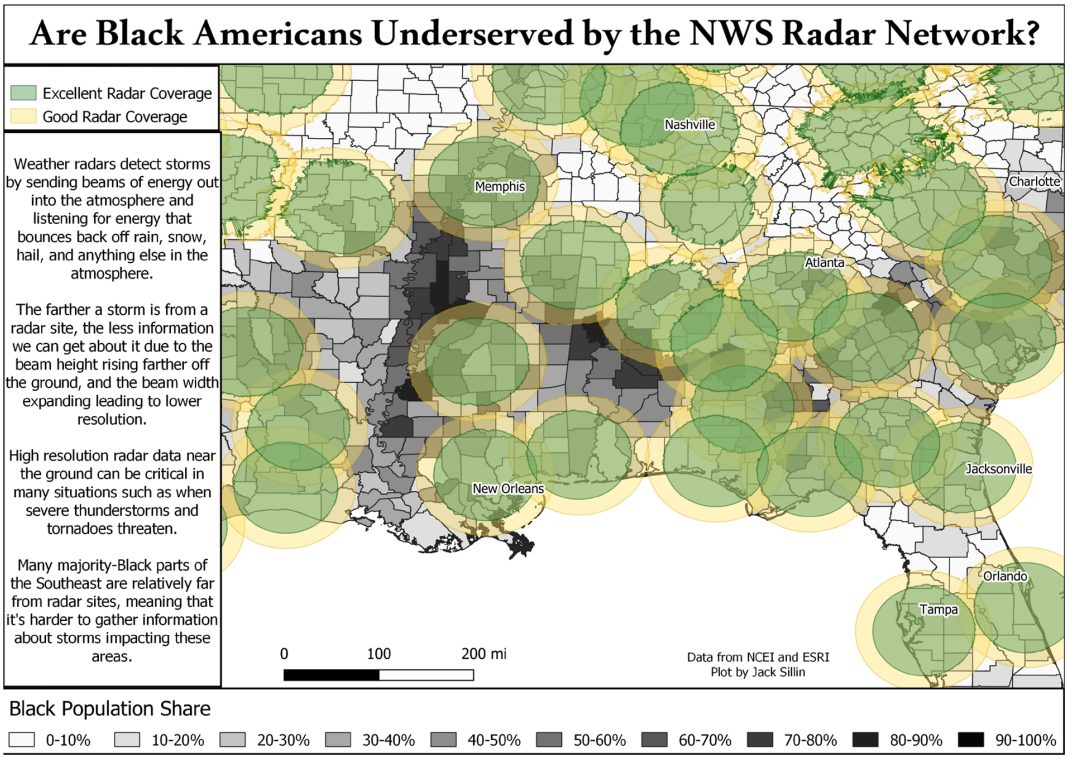
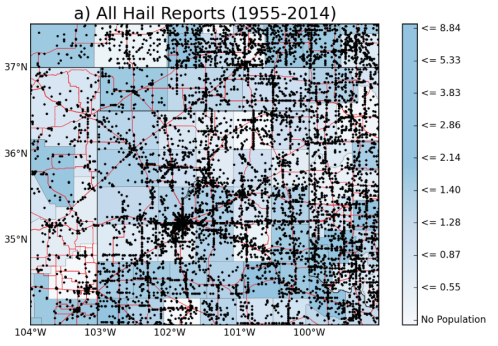
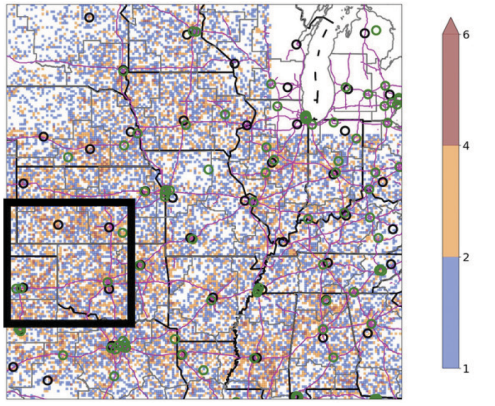


Figure 1. *Coverage of the national Doppler weather network (green and yellow circles) overlaid with the black population in the southeast United States, courtesy of Jack Sillin. This is an example of nonrepresentative data (Section 2.1.1).*

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

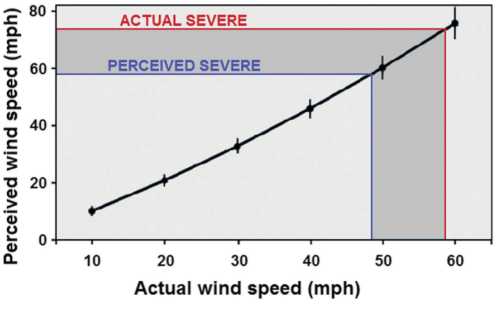


(a) Hail reports follow population

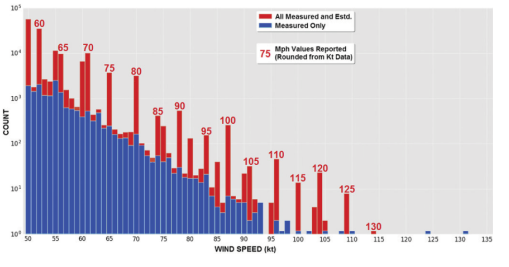


(b) Tornado reports follow population

Figure 2. *Hail and tornado reports both show a clear population bias with reports occurring more frequently along roads and cities. This can be seen as examples of (i) data being missing in low population areas (Section 2.1.1) and (ii) faulty labels (Section 2.1.2). Panel a is from Allen and Tippett (2015) © Allen and Tippett. Reproduced under the terms of the Creative Commons Attribution–NonDerivative License (CC BY-ND) Panel b is from Potvin et al. (2019) © American Meteorological Society. Used with permission. Contact copyright holder for further re-use.*



(a) Human estimates and observations of wind speed



(b) Human reports of max wind speed versus measured max wind speed from Edwards et al. (2018).

Figure 3. *Human reports of wind speed show biases in both perception of the speed itself (Panel a) and in the binning of the data (Panel b). Both panels are from Edwards et al. (2018). This example highlights both non-represenative training data (Section 2.1.1) and biased and faulty labels (Section 2.1.2). Panels 2a and b © American Meteorological Society. Used with permission. Contact copyright holder for further re-use.*

A related issue is that even if a diverse geographical area is represented in the dataset, the AI model will focus on patterns that are associated with the most common regimes and areas where the hazard occurs most frequently. It also may ignore patterns that affect a small portion of the dataset since the contribution to the overall training error is relatively small. For example, solar irradiance forecasts from an AI model trained across multiple sites in Oklahoma had higher errors at the sites in the Panhandle compared with central Oklahoma due to limited representation of Panhandle weather in the dataset (Gagne et al., 2017).

Figure 3 illustrates the typical biases of humans to estimate wind speed due to two effects. Figure 3a shows that humans tend to overestimate wind speed, and thus tend to classify it as severe at levels that are not actually considered severe. Thus, human reported wind speed results in training data with significant

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

bias toward higher wind speeds. Figure 3b illustrates that humans tend to assign wind speeds in discrete increments, with a strong bias toward multiples of five, that id, 60, 65, 70, …, kt. This example also demonstrates how nonrepresentative training data and the biases could result in an AI algorithm that does not predict true wind speed impacts correctly.

Sensor generated datasets: A potential response to these data issue could be to use only sensor-based data, such as radar-derived hail estimates. We often assume that sensors are inherently objective yet that is not guaranteed. First, the data from these sensors must be interpreted to create a full dataset. For radar-derived hail data, there are well-known overprediction biases to the data (e.g., Murillo and Homeyer, 2019). Second, data from sensors may also be incomplete, either due to sensor errors or missing sensors. With many sensors, there are areas which are well represented and areas where data are lacking. Sensor placement often depends on geological features, for example, there might be a lack of sensors at remote mountain tops or in empty fields. Economical reasons can also come into play in the placement of sensors, such as the placement of crowd-sourced sensors where there are more affluent people. Consider for example the deployment of air quality sensors. While the EPA has a network of just over 1,100 air quality sensors that are relatively evenly distributed throughout the United States, this is not the case for the popular new generation of low-cost air quality sensors, called PurpleAir. There are now over 5,500 PurpleAir sensors deployed in the United States, but they are deployed in significantly Whiter, higher income areas than the national average (deSouza and Kinney, 2021). Access to these sensors in less-privileged communities is needed to democratize air pollution data (deSouza and Kinney, 2021).

2.1.3. Data are affected by adversaries

Adversarial data are a well-known problem in machine learning (e.g., Diochnos et al., 2018; Goodfellow et al., 2018; Nelson et al., 2010). Adversaries can cause an AI model to either learn a faulty model or to be used incorrectly when applied to real-world data. For example, there are well-known examples of minor changes to speed signs causing a computer vision system to see a 35-mph speed limit as an 85-mph speed limit, something that could be quite dangerous in an autonomous car.

AI applied to environmental science data has to contend with two kinds of adversaries: humans and the environment. One needs to be aware of both types of adversaries when training and applying the AI models to the real-world.

Human adversaries: When using human reported data, there is always the possibility of users intentionally adding bad data. Figure 4 provides two examples. Figure 4a provides the screenshot of wind reported by someone who hacked the crowd-sourced mPing system for the state of Alabama. Note the long line of wind reported outlining the state border which clearly represents an intentional input of incorrect information. Such data would be detrimental to any AI model trained on these data. A second source of human adversarial data are insurance fraud (Davila et al., 2005). As the number of expensive natural disasters are increasing (Figure 4b), the number of fraudulent weather-related insurance claims is increasing, with estimates of fraudulent reports hovering around 10% (Sorge, 2018). Such fraudulent reports can also affect training data for AI by appearing in the databases of severe weather reports.

Weather as an adversary: Weather can act as its own adversary, especially when collecting observations for extreme events. For example, when seeking to observe the severity of storms, sensors can fail due to power outages caused by a strong storm (Figure 5a) or a strong storm can even destroy sensors (Figure 5b).

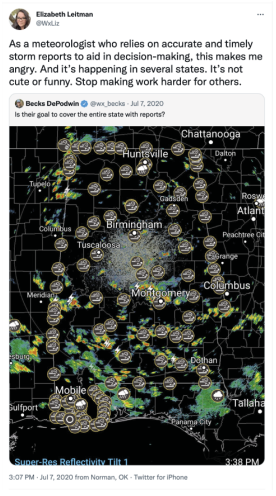
2.2. Issues related to the AI model

In addition to problems caused by bias or other complications in the training data, AI models can also develop issues on their own. Examples are given in this section.

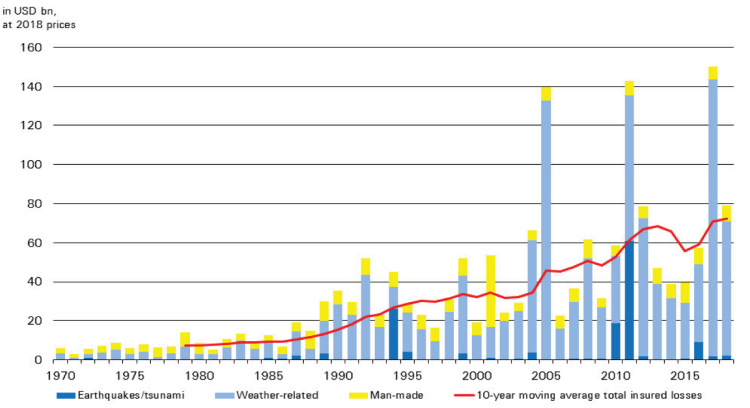
2.2.1. Model training choices

Model training choices will affect every aspect of the AI model. While AI methods are based on mathematical equations and thus often seen as "objective," there are countless choices that a scientist

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

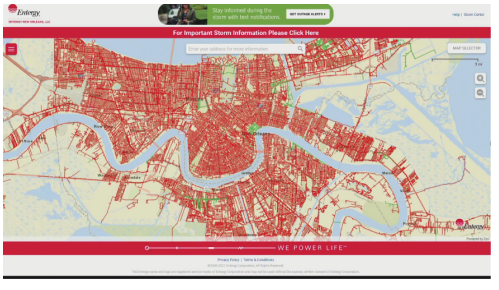


(a) Screenshot of hacked data from mPing from Cappucci (2020)



(b) Catastrophe related insured losses from 1970 to 2018 from (RE, 2018; Scotti, 2019). With at least 10% of weather-based insurance claims estimated as fraudulent (Sorge, 2018), there is increasing motivation for humans to report incorrect severe weather data

Figure 4. *Humans sometimes create adversarial data, which may be ingested by an artificial intelligence (AI) model (Section 2.1.3). While in example (a) the user intent is to create false inputs, in example, (b) the motivation for the false reports is more for personal financial gain (insurance fraud). Nevertheless, both can cause problems for AI by directly affecting reports databases used by AI models.*



(a) Power outages following a hurricane

(b) A destroyed Oklahoma Mesonet station following a severe wind gust

Figure 5. *Weather also creates its own adversaries (Section 2.1.3) by creating power outages (a) or destroying sensors (b), especially during major events. Graphics from Mesonet (2020) and Staff (2021).*

has to make that greatly affect the results. These choices include: (a) which attributes (e.g., which environmental variables) to include in the model; (b) which data sources to use; (c) how to preprocess the data (e.g., normalizing the data, removing seasonality, and applying dimension reduction methods); (d) which type of AI model to use (e.g., clustering, random forest, neural networks, etc.); and (e) how to choose hyper parameters (e.g., for random forest—how many trees, maximal depth, minimal leaf size, etc). Each of these choices has significant impact and can lead to vastly different results with severe consequences. For example, the choice of spatial resolution for the output of a model can be crucial for environmental justice. Training an AI model to predict urban heat at a low spatial resolution may average out, and thus overlook, extreme values in small neighborhoods, while using a higher spatial resolution could reveal those peaks but potentially introduce noise.

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

2.2.2. Algorithm learns faulty strategies

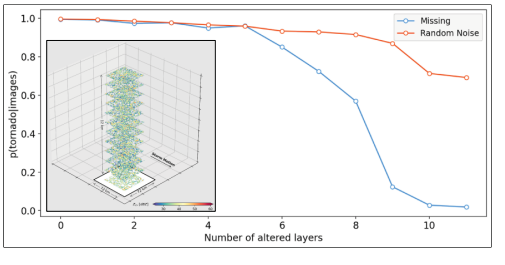
AI models are tasked to learn patterns in the data that help them come up with good estimates or classifications. Sometimes, even if the data are not faulty in the traditional sense, they may contain spurious correlations that are not representative of the real world and that the AI model learns to exploit. This issue could be discussed in the section on data-related issues (Section 2.1), but we chose to include it in this section since it is so closely coupled with AI model development, and often only diagnosed once an AI model has been developed.

One approach to identifying when the AI has learned a faulty strategy is to either use interpretable or explainable AI (XAI) (e.g., Ras et al., 2018; Molnar, 2018; McGovern et al., 2019; Samek et al., 2019; Ebert-Uphoff and Hilburn, 2020). In the case of interpretable AI, the models are intended to be understandable by human experts (Rudin, 2019) while explainable AI attempts to peer inside the black box of more complicated models and identify the internal strategies used by the models. Both approaches allow a human expert to examine the model and potentially to identify faulty strategies before deploying a model.

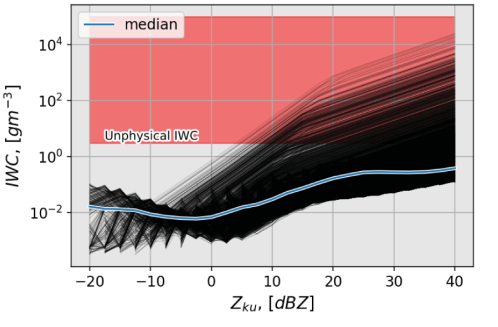
Figure 6a is an example of how an AI model can learn a faulty strategy and how this can be discovered through XAI. In this example, a deep learning model was trained to predict tornadoes (Lagerquist et al., 2020) and then examined to identify what the model had learned. Unexpectedly, we discovered that the model was still predicting a high probability of tornadoes even if the reflectivity image contained mostly noise (Chase and McGovern, 2022).

2.2.3. AI learns to fake something plausible

The emergence of an advanced type of AI algorithm, namely the generative adversarial network (GAN) introduced by Goodfellow et al. (2014), yields the ability to generate imagery that looks extremely realistic. The result of a GAN can be seen as one possible solution (an ensemble member) which might not be representative of the distribution of possible solutions. An important question is how forecasters would respond to GAN imagery with intricate small scale features that look overwhelmingly realistic, but might not be as accurate as they appear to be. Ravuri et al. (2021) conducted a cognitive assessment study with expert meteorologists to evaluate how forecasters respond to GAN imagery of radar for precipitation nowcasting. They concluded that for their application *forecasters made deliberate judgements of the predictions by relying on their expertise, rather than being swayed by realistic looking images* (Suppl.



(a) Tornado prediction model from Lagerquist et al. (2020) and Randy Chase (Chase and McGovern, 2022). ©Chase and McGovern. Used with permission. Contact copyright holder for further re-use.



(b) Example of a partial dependency plot (Molnar, 2018) which shows how an AI model could product unphysical results when given unexpected inputs (Section 2.2.4). From (Chase et al., 2021). ©Molnar. Reproduced under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC-BY-NC-SA 4.0).

Figure 6. *Examples of an AI learning to fake something plausible (Section 2.2.3) and of a model generating unexpected output when applied to data outside of the range of expected data (Section 2.2.4).*

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

C.6. of Ravuri et al., 2021). Nevertheless, the jury is still out whether that is generally the case. Such methods might sometimes trick forecasters into greater trust than is warranted. Similar methods are used to generate deep fake images and videos (Adee, 2020) which have successfully fooled people into believing misinformation (Vaccari and Chadwick, 2020). If similar approaches are used in environmental science, trust in AI models could be lost.

One of the major dangers of generative models used in contexts where they are parameterizing fine scale data (Gagne et al., 2020) or filling in missing data (Geiss and Hardin, 2021) is the lack of guarantees about what data they use to fill in the gaps. If an ML model is trained without any physical constraints, it is possible for the algorithm to insert spurious weather features. If this occurs in otherwise data sparse regions, forecasters may not have any additional data to discredit the false prediction for many hours. More subtle discrepancies would be the most problematic since they are harder for forecasters to catch.

2.2.4. Models used in inappropriate situations

Another model-based issue that can arise is when an AI model is trained for one situation, perhaps a specific weather regime or for a specific geographical area, and then applied to new situations where it is not physically meaningful. In such situations, the AI may predict incorrectly, without giving any warning, which could lead to loss of life and environmental injustice.

For example, Figure 6b shows a partial dependence plot (Molnar, 2018) of the predictions of an AI method predicting total ice water content (i.e., mass of snow in a cloud) from radar data (Chase et al., 2021). If input given to the model is uncharacteristic of the training dataset, the output may not be physically realistic. This issue also comes up frequently when ML is used to replace numerical weather prediction parameterization schemes. If the ML model is transferred to a different climate or geographic regime or even to the hemisphere opposite where it was trained (where the sign of many common variables are flipped), it often performs quite poorly.

2.2.5. Nontrustworthy models deployed

Ensuring that a model is trustworthy is a key part of ethical and responsible AI. Often AI systems are *deployed* before they are ready, in particular before they have been thoroughly tested and validated by domain experts. Figure 7 shows the results of a commercial AI system predicting damage caused by a hypothetical 7.0-magnitude earthquake in the Seattle area. The three panels show results from three different version of the system delivered to the city of Seattle to assist them to plan where to establish earthquake shelters. The widely differing results in the three panels arose from using different data sources, but also from programming errors (e.g., counting each apartment in a high-rise building as a separate structure). Deploying the model before it was ready for operation forced the city to revamp nearly

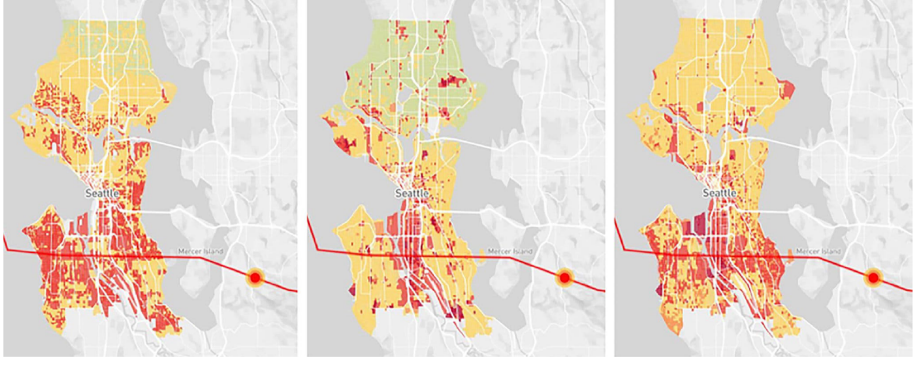


Figure 7. *Three vastly different damage predictions for the same hypothetical 7.0-magnitude Seattle area earthquake delivered by different versions of the same AI system. Figure from Fink (2019), crediting the Seattle Office of Emergency Management.*

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

completed plans for sheltering earthquake-displaced residents that were developed using the original version (Fink, 2019). The trust in the AI system was so eroded that the city considered terminating their contract, and only kept it because the company found an external funder and could offer it to Seattle for free.

Typical mistakes to be avoided, illustrated by this example, are developing a system without engaging domain experts or local end-users, deploying a system before it has been rigorously validated, and over-promising the system's accuracy to the end user without any scientific evidence. For the latter, there are many new methods for uncertainty quantification in machine learning models that we as a community should start utilizing routinely for this purpose (e.g., Barnes et al., 2021; Orescanin et al., 2021).

2.2.6. Lack of robustness

While a lack of robustness could also be categorized under models being inappropriately applied, adversarial data, or even a lack of trust, we separate it into its own category to highlight the need for robustness in ethical and responsible AI. Robust AI for environmental sciences will fail gracefully when outputs are out of range (e.g., Section 2.2.4) and will also constrain the outputs by the laws of physics. For instance, when preparing a workshop tutorial, one of the authors was demonstrating a simple ML method that could predict a local temperature. When presented with an input that was two standard deviations outside of its training data, the method suddenly predicted a temperature hotter than the surface of the sun. Such predictions may contribute to a lack of trust and create environmental injustices. Furthermore, integrating advances in AI methods that enable physics and/or fairness principles (Kearns and Roth, 2019; Thomas et al., 2019; Mehrabi et al., 2021) into the learning methods could both align with and help advance ethical and responsible AI.

2.3. Other issues related to workforce and society

We have discussed ways that both the training data and the AI model could go wrong. Here, we focus on the broader picture of how AI is applied and developed and how that relates to environmental justice. This aspect also needs to be addressed in developing ethical and responsible AI for environmental sciences.

2.3.1. Globally applicable AI approaches may stymie burgeoning efforts in developing countries

Large organizations developing and deploying AI systems, including government agencies and private companies, generally want their AI systems to be globally applicable. People in countries outside that of the provider organization can benefit from technological advances that may not otherwise be available to them. Provider organizations can benefit from a broader set of customers and stakeholders. These kinds of global activities work best when researchers and companies directly engage with different local populations and incorporate their needs and concerns into the design process. For example, Google's crisis team has been working with multiple local and government partners in India and Bangladesh to create a hyper-local flood forecasting product (Matias, 2021).

However, the relative ease of deploying AI on global datasets can also increase the temptation for organizations to automate deployment without fully monitoring its impact. Without direct engagement with various populations, organizations may not be aware of the unintended consequences of their algorithms. They also may not have access to observations in these areas to validate their performance. They may also undercut local providers who are more aligned with the community but do not have the same resources and economies of scale to compete on price, especially if the service is offered for free or is supported by advertising revenue. Government meteorological centers in both developed and developing countries have dealt with budget cuts and cost-recovery requirements even before the growth of AI and the private weather industry, so further adoption of AI could exacerbate these trends.

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

For example, mobile weather apps have become the primary source of weather information for younger people (Phan et al., 2018). However, the apps' uniform presentation of weather information may result in users not being aware of more localized and rare weather hazards. A local government or television meteorologist on the other hand can tailor their communication to highlight these issues.

2.3.2. *Lack of input or consent on data collection and model training*

As illustrated in the discussion of Figure 7, when collecting local data for training an AI model or when creating an AI model to inform decisions in or about a specific place, it is critical that the people affected and those with local expertise are involved from the beginning (Renn et al., 1995; Stern and Fineberg, 1996; Chilvers, 2009; Voinov et al., 2016; Pidgeon, 2021). This is more typically known in the social sciences but not nearly as common in AI.

The types of data that are collected affect the AI models that can be applied and trained. Understanding that data provide a source of power and that what is not collected cannot be used (D'Ignazio and Klein, 2020), it is possible to see that AI could too easily be used to perpetuate environmental and climate injustices if data are not collected carefully.

In environmental research it has long been noted that engaging those whose concerns the science is purportedly informing or addressing can improve models and outcomes, is a moral obligation, and is the virtuous course of action (e.g., Lemos and Morehouse, 2005). A core duty in environmental sciences, as in all sciences, is to be honest (e.g., National Academies of Sciences, Engineering, and Medicine (NASEM), 2009; 2019), which respects the dignity and rights of persons (e.g., Keohane et al., 2014). In AI for environmental sciences this requires grappling with how to be transparent (i.e., show your work—D'Ignazio and Klein, 2020) about methodological biases and uncertainties that are often complex and poorly understood. When scientific conditions such as sensitivity to training conditions or data provenance are not reported, that lack of transparency both disempowers and disrespects those the model may be intended to help, and can make a study irreplicable and even irreproducible (NASEM, 2019).

It is thus critical to keep in mind that many types of data and knowledge already exist outside of the colonial framework in which most AI data have been collected. For example, scientists must work directly with local leaders to incorporate Indigenous knowledge (Hiwasaki et al., 2014; Cartier, 2019) in an ethical and responsible manner.

2.3.3. *Scientists feeling disenfranchised*

There are two primary reasons why scientists might feel disenfranchised.

Scientists not fluent in AI might feel disenfranchised: Environmental scientists are absolutely essential for the development of meaningful AI tools for the their domains. Currently, with a lack of education in AI for environmental sciences, leading scientists may feel excluded from advancing state-of-the-art tools. They must either learn about AI themselves or collaborate with AI scientists to be part of major future development in their field. As a consequence they may feel less independent and less appreciated. Clearly, we need more educational resources to help environmental scientists learn the basics of AI in order to empower them to play a leading role in future developments.

Required computational resources might limit research opportunities to privileged groups: While AI methods for many environmental science applications require only modest computational resources, some research now requires access to extensive high performance computing facilities, which limits this kind of research to those research groups privileged enough to have such access. To level the playing field we as a community need to find ways to develop an open science model that provides access to computational resources also for other groups who need them (e.g., Gentemann et al., 2021).

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

2.3.4. Increase of CO_2 emissions due to computing

It is well known that the increasing computational demands of AI training are responsible for a surprisingly large carbon footprint worldwide (Schwartz et al., 2020; Xu et al., 2021). Thus, we need to consider, and limit, the carbon footprint from the computational needs of environmental science. Green deep learning (aka Green AI) is a new research field that appeals to researchers to pay attention to the carbon footprint of their research, and to focus on using lightweight and efficient methods (Xu et al., 2021). We should fully explore these new developments, including focusing on how to make ML models simpler for environmental science applications, which would have many advantages, including increased transparency and robustness (Rudin, 2019), but also lower computational needs and lower carbon footprints.

3. Discussion and Future Work

The scientific community is still grappling with the many ethical questions raised by the introduction of AI in general (O'Neil, 2016; Floridi, 2019; Schmidt et al., 2021) and for Earth science applications in particular (Doorn, 2021; Coeckelbergh, 2021). The effort presented here represents the beginning of our work on developing a full understanding of the need for ethical, responsible, and trustworthy AI for the environmental sciences and of the interactions between ethical and responsible AI and trustworthy AI within the *NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES, ai2es.org)*. In an upcoming paper, we plan to present guiding principles for ethical, responsible, and trustworthy AI for the environmental sciences.

Acknowledgments. We are grateful for the graphic on radar coverage in the southeast US contributed by Jack Sillin (Cornell), the graphics contributed by Randy Chase (OU), and references contributed by Amy Burzynski (CSU).

Ethics Statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Data Availability Statement. This manuscript does not develop any new data or code.

Author Contributions. Conceptualization: A.M., I.E.-U., D.J.G., and A.B.; Data visualization: A.M. and I.E.; Methodology: A.M., I.E.-U., D.J.G., and A.B.; Writing original draft: A.M. and I.E equally shared the majority of the writing of this paper. All authors contributed to the writing and approved the final submitted draft.

Funding Statement. This material is based upon work supported by the National Science Foundation under Grant No. ICER-2019758 and by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Competing Interests. The authors declare no competing interests exist.

References

Adee S (2020) *What are 6 deepfakes and how are they created? IEEE Spectrum.* Available at https://spectrum.ieee.org/what-is-deepfake. Accessed Nov 18 2022

Allen J and Tippett M (2015) The characteristics of United States hail reports: 1955–2014. *Electronic Journal of Severe Storms Meteorology 10*, 1–31.

Barnes EA, **Barnes RJ and Gordillo N** (2021) Adding uncertainty to neural network regression tasks in the geosciences. Preprint, arXiv:2109.07250.

Benjamin R (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity.

Cappucci M (2020) *NOAA storm-spotting app was suspended after being overrun with false and hateful reports. The Washington Post.* Available at https://www.washingtonpost.com/weather/2020/07/14/noaa-app-mping-suspended/.

Cartier KMS (2019) Keeping indigenous science knowledge out of a colonial mold. EOS Science News by AGU. Available at https://eos.org/articles/keeping-indigenous-science-knowledge-out-of-a-colonial-mold.

Chase R and McGovern A (2022) Deep learning parameter considerations when using radar and satellite measurements. In *21st Conference on Artificial Intelligence for Environmental Science at the 102nd American Meteorological Society Annual Meeting*. Houston, TX: American Meteorological Society.

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

Chase RJ, **Nesbitt SW and McFarquhar GM** (2021) A dual-frequency radar retrieval of two parameters of the snowfall particle size distribution using a neural network. *Journal of Applied Meteorology and Climatology 60*(3), 341–359. https://doi.org/10.1175/JAMC-D-20-0177.1. Available at https://journals.ametsoc.org/view/journals/apme/60/3/JAMC-D-20-0177.1.xml.

Chilson C, **Avery K**, **McGovern A**, **Bridge E**, **Sheldon D and Kelly J** (2019) Automated detection of bird roosts using NEXRAD radar data and convolutional neural networks. *Remote Sensing in Ecology and Conservation 5*(1), 20–32. https://doi.org/10.1002/rse2.92. Available at https://onlinelibrary.wiley.com/doi/10.1002/rse2.92.

Chilvers J (2009) Deliberative and participatory approaches in environmental geography. In *A Companion to Environmental Geography*. Oxford, UK: Wiley-Blackwell, 400–417.

Coeckelbergh M (2021) Ai for climate: freedom, justice, and other ethical and political challenges. *AI and Ethics 1*(1), 67–72.

D'Ignazio C and Klein LF (2020) *Data Feminism*. Cambridge, MA: MIT Press.

Davila M, **Marquart JW and Mullings JL** (2005) Beyond mother nature: contractor fraud in the wake of natural disasters. *Deviant Behavior 26*(3), 271–293. https://doi.org/10.1080/01639620590927623.

deSouza P and Kinney PL (2021) On the distribution of low-cost pm 2.5 sensors in the us: demographic and air quality associations. *Journal of Exposure Science & Environmental Epidemiology 31*(3), 514–524.

Diochnos D, **Mahloujifar S and Mahmoody M** (2018) Adversarial risk and robustness: general definitions and implications for the uniform distribution. In Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N and Garnett R (eds), *Advances in Neural Information Processing Systems 31*. Red Hook, NY: Curran Associates, pp. 10359–10368. Available at http://papers.nips.cc/paper/8237-adversarial-risk-and-robustness-general-definitions-and-implications-for-the-uniform-distribution.pdf.

Doorn N (2021) Artificial intelligence in the water domain: opportunities for responsible use. *Science of the Total Environment 755*, 142–561.

Duporge I, **Isupova O**, **Reece S**, **Macdonald DW and Wang T** (2021) Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. *Remote Sensing in Ecology and Conservation 7*(3), 369–381. https://doi.org/10.1002/rse2.195. https://onlinelibrary.wiley.com/doi/10.1002/rse2.195.

Ebert-Uphoff I and Hilburn K (2020) Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. *Bulletin of the American Meteorological Society 101*(12), E2149–E2170.

Edwards R, **Allen JT and Carbin GW** (2018) Reliability and climatological impacts of convective wind estimations. *Journal of Applied Meteorology and Climatology 57*(8), 1825–1845. https://doi.org/10.1175/JAMC-D-17-0306.1. Available at https://journals.ametsoc.org/view/journals/apme/57/8/jamc-d-17-0306.1.xml.

EPA (2022) Learn about environmental justice. Available at https://www.epa.gov/environmentaljustice/learn-about-environmental-justice.

Fink S (2019) This high-tech solution to disaster response may be too good to be true. New York Times. Available at https://www.nytimes.com/2019/08/09/us/emergency-response-disaster-technology.html (accessed 12 August 2020).

Floridi L (2019) Establishing the rules for building trustworthy AI. *Nature Machine Intelligence 1*(6), 261–262.

Gagne DJ, **Christensen HM**, **Subramanian AC and Monahan AH** (2020) Machine learning for stochastic parameterization: generative adversarial networks in the Lorenz'96 model. *Journal of Advances in Modeling Earth Systems 12*(3), e2019MS001896. Available at https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001896.

Gagne DJ, **McGovern A**, **Haupt SE and Williams JK** (2017) Evaluation of statistical learning configurations for gridded solar irradiance forecasting. *Solar Energy 150*, 383–393. https://doi.org/10.1016/j.solener.2017.04.031. Available at https://www.sciencedirect.com/science/article/pii/S0038092X17303158.

Geiss A and Hardin JC (2021) Inpainting radar missing data regions with deep learning. *Atmospheric Measurement Techniques 14*(12), 7729–7747. https://doi.org/10.5194/amt-14-7729-2021. Available at https://amt.copernicus.org/articles/14/7729/2021/.

Gensini VA, **Walker CC**, **Ashley S and Taszarek M** (2021) Machine learning classification of significant tornadoes and hail in the United States using era5 proximity soundings. *Weather and Forecasting 36*(6), 2143–2160. https://doi.org/10.1175/WAF-D-21-0056.1. Available at https://journals.ametsoc.org/view/journals/wefo/36/6/WAF-D-21-0056.1.xml.

Gentemann CL, **Holdgraf C**, **Abernathey R**, **Crichton D**, **Colliander J**, **Kearns EJ**, **Panda Y and Signell RP** (2021) Science storms the cloud. *AGU Advances 2*(2), e2020AV000354. https://doi.org/10.1029/2020AV000354.

Goodfellow I, **Pouget-Abadie J**, **Mirza M**, **Xu B**, **Warde-Farley D**, **Ozair S**, **Courville A and Bengio Y** (2014) Generative adversarial nets. In *Advances In Neural Information Processing Systems*, Proceedings of the 27th International Conference on Neural Information Processing Systems pp.2672–2680.

Goodfellow IJ, **McDaniel PD and Papernot N** (2018) Making machine learning robust against adversarial inputs. *Communications of the ACM 61*(7), 56–66.

Hilburn KA, **Ebert-Uphoff I and Miller SD** (2021) Development and interpretation of a neural-network-based synthetic radar reflectivity estimator using GOES-R satellite observations. *Journal of Applied Meteorology and Climatology 60*(1), 3–21.

Hill AJ and Schumacher RS (2021) Forecasting excessive rainfall with random forests and a deterministic convection-allowing model. *Weather and Forecasting 36*(5), 1693–1711. https://doi.org/10.1175/WAF-D-21-0026.1. Available at https://journals.ametsoc.org/view/journals/wefo/36/5/WAF-D-21-0026.1.xml.

Hiwasaki L, **Luna E**, **Syamsidik S and Shaw R** (2014) Process for integrating local and indigenous knowledge with science for hydro-meteorological disaster risk reduction and climate change adaptation in coastal and small island communities. *International Journal of Disaster Risk Reduction 10*, 15–27.

Kantayya, S (2020) Film: coded bias, documentary, A Shalini Kantayya film, 7th Empire Media.

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

Karpatne A, **Jiang Z**, **Vatsavai RR**, **Shekhar S and Kumar V** (2016) Monitoring land-cover changes: a machine-learning perspective. *IEEE Geoscience and Remote Sensing Magazine 4*(2), 8–21. https://doi.org/10.1109/MGRS.2016.2528038.

Kearns M and Roth A (2019) *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. New York: Oxford University Press.

Keohane RO, **Lane M and Oppenheimer M** (2014) The ethics of scientific communication under uncertainty. *Politics, Philosophy & Economics 13*(4), 343–368. https://doi.org/10.1177/1470594X14538570.

Lagerquist R, **Stewart JQ**, **Ebert-Uphoff I and Kumler C** (2021) Using deep learning to now cast the spatial coverage of convection from Himawari-8 satellite data. *Monthly Weather Review 149*(12), 3897–3921. https://doi.org/10.1175/MWR-D-21-0096.1. Available at https://journals.ametsoc.org/view/journals/mwre/149/12/MWR-D-21-0096.1.xml.

Lagerquist RA, **McGovern CR**, **Homeyer DJG II and Smith T** (2020) Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Monthly Weather Review 148* (7), 2837–2861. https://doi.org/10.1175/MWR-D-19-0372.1. Available at https://journals.ametsoc.org/view/journals/mwre/148/7/mwrD190372.xml.

Lemos MC and Morehouse BJ (2005) The co-production of science and policy in integrated climate assessments. *Global Environmental Change 15*(1), 57–68.

Lin TY, **Winner K**, **Bernstein G**, **Mittal A**, **Dokter AM**, **Horton KG**, **Nilsson C**, **Van Doren BM**, **Farnsworth A**, **La Sorte FA and Maji S** (2019) MistNet: measuring historical bird migration in the US using archived weather radar data and convolutional neural networks. *Methods in Ecology and Evolution 10*(11), 1908–1922. https://doi.org/10.1111/2041-210x.13280. Available at https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.13280.

Matias Y (2021) Expanding our ML-based flood forecasting. Available at https://blog.google/technology/ai/expanding-our-ml-based-flood-forecasting/.

McGovern A, **Lagerquist R**, **Gagne D**, **Jergensen G**, **Elmore K**, **Homeyer C and Smith T** (2019) Making the black box more transparent: understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society 100* (11), 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1.

Mehrabi N, **Morstatter F**, **Saxena N**, **Lerman K and Galstyan A** (2021) A survey on bias and fairness in machine learning. *ACM Computing Surveys 54*(6). https://doi.org/10.1145/3457607.

Mesonet O (2020) Top 20 extreme weather events in mesonet history. Available at https://www.mesonet.org/20th/.

Mithal V, **Nayak G**, **Khandelwal A**, **Kumar V**, **Nemani R and Oza NC** (2018) Mapping burned areas in tropical forests using a novel machine learning framework. *Remote Sensing 10*(1), 69. https://doi.org/10.3390/rs10010069. Available at https://www.mdpi.com/2072-4292/10/1/69.

Molnar C (2018) *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Victoria, BC: Leanpub. Available at https://christophm.github.io/interpretable-ml-book/.

Murillo EM and Homeyer CR (2019) Severe hail fall and hailstorm detection using remote sensing observations. *Journal of Applied Meteorology and Climatology 58*(5), 947–970. https://doi.org/10.1175/JAMC-D-18-0247.1. Available at https://journals.ametsoc.org/view/journals/apme/58/5/jamc-d-18-0247.1.xml.

National Academies of Sciences, Engineering, and Medicine (NASEM) (2009) *On Being a Scientist: A Guide to Responsible Conduct in Research*. Washington, DC: National Academies Press.

National Academies of Sciences, Engineering, and Medicine (NASEM) (2019) *Reproducibility and Replicability in Science*. Washington, DC: National Academies Press.

Nelson B, **Rubinstein BIP**, **Huang L**, **Joseph AD and Tygar JD** (2010). Classifier evasion: Models and open problems. In *International Workshop on Privacy and Security Issues in Data Mining and Machine Learning* (pp. 92–98). Springer, Berlin, Heidelberg.

O'Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.

Orescanin M, **Petković V**, **Powell SW**, **Marsh BR and Heslin SC** (2021) Bayesian deep learning for passive microwave precipitation type detection. *IEEE Geoscience and Remote Sensing Letters*, *19*, pp.1–5.

Phan MD, **Montz BE**, **Curtis S and Rickenbach TM** (2018) Weather on the go: an assessment of smartphone mobile weather application use among college students. *Bulletin of the American Meteorological Society 99*(11), 2245–2257. https://doi.org/10.1175/BAMS-D-18-0020.1. Available at https://journals.ametsoc.org/view/journals/bams/99/11/bams-d-18-0020.1.xml?tab_body=pdf.

Pidgeon N (2021) Engaging publics about environmental and technology risks: frames, values and deliberation. *Journal of Risk Research 24*(1), 28–46. https://doi.org/10.1080/13669877.2020.1749118.

Potvin CK, **Broyles C**, **Skinner PS**, **Brooks HE and Rasmussen E** (2019) A Bayesian hierarchical modeling framework for correcting reporting bias in the U.S. tornado database. *Weather and Forecasting 34*, 15–30. https://doi.org/10.1175/WAF-D-18-0137.1.

Ras G, **van Gerven M and Haselager P** (2018) Explanation methods in deep learning: users, values, concerns and challenges. In *Explainable and Interpretable Models in Computer Vision and Machine Learning*. Cham: Springer, pp. 19–36.

Ravuri S, **Lenc K**, **Willson M**, **Kangin D**, **Lam R**, **Mirowski P**, **Fitzsimons M**, **Athanassiadou M**, **Kashem S**, **Madge S and Prudden R**, (2021) Skilful precipitation nowcasting using deep generative models of radar. *Nature 597*(7878), 672–677.

Re S (2018) Preliminary sigma estimates for 2018: global insured losses of usd 79 billion are fourth highest on sigma records. Available at https://www.swissre.com/media/news-releases/nr_20181218_sigma_estimates_for_2018.html.

https://doi.org/10.1017/eds.2022.5 Published online by Cambridge University Press

Reichstein M, **Camps-Valls G**, **Stevens B**, **Jung M**, **Denzler J**, **Carvalhais N and Prabhat** (2019) Deep learning and process understanding for data-driven Earth system science. *Nature 566*, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Renn O, **Webler T and Wiedemann P** (1995) *Fairness and Competence in Citizen Participation: Evaluating Models for Environmental Discourse*. Dordrecht: Kluwer Academic Press.

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence 1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x.

Samek W, **Montavon G**, **Vedaldi A**, **Hansen L and Muller K-R** (eds.) (2019) *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Berlin: Springer Nature.

Schmidt E, **Work B**, **Catz S**, **Chien S**, **Darby C**, **Ford K**, **Griffiths JM**, **Horvitz E**, **Jassy A**, **Mark W and Matheny J** (2021) *National Security Commission On Artificial Intelligence (AI).* Technical Report. National Security Commission on Artificial Intelligence.

Schumacher RS, **Hill AJ**, **Klein M**, **Nelson JA**, **Erickson MJ**, **Trojniak SM and Herman GR** (2021) From random forests to flood forecasts: a research to operations success story. *Bulletin of the American Meteorological Society 102*(9), E1742–E1755. https://doi.org/10.1175/BAMS-D-20-0186.1. https://journals.ametsoc.org/view/journals/bams/102/9/BAMS-D-20-0186.1.xml.

Schwartz R, **Dodge J**, **Smith NA and Etzioni O** (2020) Green AI. *Communications of the ACM 63*(12), 54–63.

Scotti V (2019) How cities can become more resilient to climate change. *World Economic Forum.* Available at https://www.weforum.org/agenda/2019/06/can-we-be-resilient-to-climate-change/.

Shepherd M (2021) Are black and rural residents in the south more vulnerable to tornadoes due to radar gaps? *Forbes*, Available at https://www.forbes.com/sites/marshallshepherd/2021/03/20/are-black-and-rural-residents-in-the-south-more-vulnerable-to-tornadoes-due-to-radar-gaps/?sh=58c1d4014988.

Sillin J (2021) Twitter post. Available at https://twitter.com/JackSillin/status/1372957704138981378?s=20.

Sorge G (2018) Weather-related events and insurance fraud. *Property Casualty 360.* Available at https://www.propertycasualty360.com/2018/10/02/weather-related-events-and-insurance-fraud/?slreturn=20211026223239.

Staff W (2021) *Recovery time unclear, check map for your parish. WWLTV.* Available at https://www.wwltv.com/article/weather/hurricane/widespread-power-outages-reported-9400-in-the-dark/289-d8a78748-9a37-4937-90af-0d2d7cb3fbd6.

Stern PC and Fineberg HC (1996) *Understanding Risk: Informing Decisions in a Democratic Society.* Washington, DC: US National Research Council.

Thomas PS, **da Silva BC**, **Barto AG**, **Giguere S**, **Brun Y and Brunskill E** (2019) Preventing undesirable behavior of intelligent machines. *Science 366*(6468), 999–1004. https://www.science.org/doi/abs/10.1126/science.aag3311.

Vaccari C and Chadwick A (2020) Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society 6*(1), 1–13.

Voinov A, **Kolagani N**, **McCall MK**, **Glynn PD**, **Kragt ME**, **Ostermann FO**, **Pierce SA and Ramu P** (2016) Modelling with stakeholders—next generation. *Environmental Modelling & Software 77*, 196–220.

Xu J, **Zhou W**, **Fu Z**, **Zhou H and Li L** (2021) A survey on green deep learning. Preprint, arXiv:2111.05193.

Cite this article: McGovern A. Ebert-Uphoff I. Gagne D. and Bostrom A. (2022). Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Environmental Data Science*, 1: e6. doi:10.1017/eds.2022.5