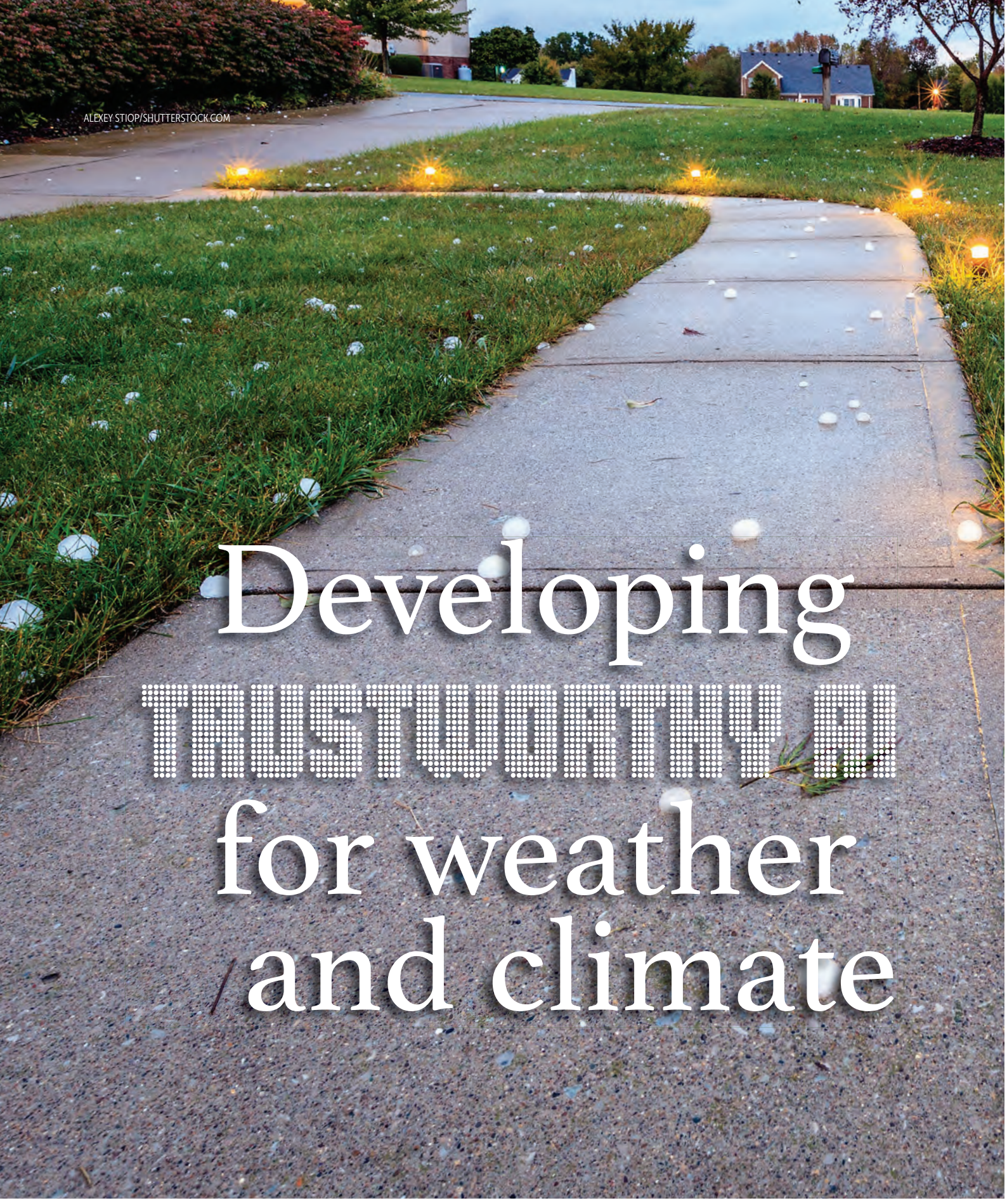


ALEXEY STIOP/SHUTTERSTOCK.COM

Developing TRUSTWORTHY AI for weather and climate

Amy McGovern directs the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES) and is a professor in computer science and meteorology at the University of Oklahoma in Norman. **Philippe Tissot** coleads the coastal oceanography team at AI2ES and is the chair for coastal artificial intelligence at Texas A&M University–Corpus Christi. **Ann Bostrom** coleads the risk communication team at AI2ES and is an environmental policy professor at the University of Washington in Seattle.

Amy McGovern, Philippe Tissot, and Ann Bostrom

By improving the prediction, understanding, and communication of powerful events in the atmosphere and ocean, artificial intelligence can revolutionize how communities respond to climate change.

The year is 2028 and the weather continues to produce climate-induced extremes, but something has changed. Your phone is now giving you early, accurate warnings to help you prepare.

Major heat wave hitting the SW United States in 3 weeks. Be prepared for an extended period of extreme temperatures and higher humidity than usual.

Warning: Baseball-sized hail and strong winds from the north are extremely likely to hit your house in approximately 20 minutes. Move belongings inside, and stay away from any north-facing windows.

Extreme cold temperatures are arriving in your area in 3 days and will last for at least 4 days. Prepare now to ensure your pipes do not freeze, and be ready for potentially extended periods of electrical outages.

Imagine that high-impact weather phenomena, such as those described above, are forecast with sufficiently advanced warning and precision that humankind is able to significantly mitigate the effects of such events globally. Furthermore, the predictions are known to be trustworthy, so individuals and local and state governments can act immediately to save lives and property.

Such a scenario is not just a vision: It may be a reality in a few years. As the climate changes, weather extremes are affecting species and ecosystems around the globe—and are becoming more extreme (see the article by Michael Wehner, Physics Today, September 2023, page 40). At the same time, recent developments in artificial intelligence (AI) and machine learning (ML) are showing how that vision might be realized.

AI offers multiple methods for handling large quantities of data, helping automate processes, and providing information to human decision makers.[1] Traditional AI methods have been used in environmental sciences for years.[2] Such methods include statistical techniques, such as linear regression, and basic object-grouping methods, such as clustering. Both have a history in environmental-science dating back several decades.[3] A little over a decade ago, weather and climate phenomena began to be understood with more-modern AI techniques, including decision trees—basically flowcharts created by an algorithm rather than constructed by hand—and groups of trees known as random forests.

ML, a subset of AI, focuses on methods that use data to learn and adapt so that they're

SEA TURTLES were rescued off the coast of Texas by volunteers in February 2022 **(left)** and January 2018 **(right)** after the successful prediction of a cold-stunning weather event by an artificial-intelligence-based forecasting model. After measurements of the turtles were taken, they were transported to a rehabilitation facility. (Courtesy of AI2ES.)

generalizable to novel situations. When AI is discussed in the news, it is most often referring to a specific form of ML called deep learning,[4] which has become popular lately. The key changes facilitating the explosion of deep learning have been the creation of innovative ways to handle spatial and temporal dependencies in the data and corresponding hardware improvements, which have made it possible for neural networks, a type of deep learning, to be trained with millions of parameters.

Deep learning has revolutionized the field of AI across various applications, including language translation, game theory, and image recognition (see, for example, the article by Sankar Das Sarma, Dong-Ling Deng, and Lu-Ming Duan, Physics Today, March 2019, page 48). AI methods can do the same for weather and climate predictions too (see reference 5 and Physics Today, May 2019, page 32). For example, multiple recent papers have introduced global weather-forecasting systems based entirely on AI methods. Although those systems need to be trained by traditional numerical weather-prediction models, their predictions are made solely through a deep-learning algorithm and do not depend on physics-based equations.[6]

Despite the long development history of AI methods for predicting weather and climate events, few have been implemented operationally by NOAA and private industry. Early operational AI models were based on relatively simple architectures, such as tree-based designs that can be read by humans. Several new startup companies and larger, established companies, however, are focused on applying more complex AI methods to commercial weather-prediction products. NOAA has also recently begun to deploy AI methods for targeted applications. With all the changes, it is critical that AI methods are beneficial to society, that they can be gauged by their users for their applicability, and that their predictions can be trusted.

Developing and deploying trustworthy AI requires a diverse multidisciplinary research team. The team at the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), for which the three of us work, consists of AI developers, social scientists, atmospheric and ocean scientists, and end users. AI2ES is rapidly developing new AI methods that will enable us to improve our scientific understanding and prediction of high-impact weather and climate phenomena, user trust in AI products, and our communication of AI's risks.[7]

Developing trustworthy AI

The diagram on page 29 outlines how the different pieces of AI2ES work together to create trustworthy AI. Traditional AI work is often done by only computer-science researchers, but our synergistic team is made up of researchers in AI, atmospheric science, coastal oceanography, and risk communication. Our goal is to ensure that we meet the needs of our end users—primarily forecasters and emergency managers—and that we understand what it means for AI to be trustworthy.

In any risky situation, successfully communicating and managing risk depends on the trust between those involved.[8] When applying AI methods to climate and extreme-weather forecasting, the uncertainties of AI need to be added to the uncertainties of the environmental predictions. The com-

pounding uncertainties raise the stakes for effectively communicating the risks and make trust even more critical. When trust in AI is low, AI-based forecasts and warnings may be ignored or misconstrued. AI, therefore, needs to be both trusted and trustworthy to be used in various high-risk situations.

Trust is usually enhanced by relevant evidence of competence and reliability,[9] but trust in an AI model is also contingent on people believing that the model aligns with their own interests. Biased or poor-quality training data can lead to biased or more-uncertain AI forecasts, which have the potential to harm those whose actions depend on the forecasts.

Models in Earth sciences are used for many purposes. Some examples at AI2ES include predicting freezes for various environmental-management purposes, protecting endangered species, and forecasting and warning for severe convective storms to protect people and save lives. Risk attitudes and trust are known to vary by the nature of the decision and the decision context[10]—who controls the decision making, for example, and how catastrophic the consequences might be—and by the attributes of the modeling system and modeling context.[11] For those reasons, understanding the nature of trust and developing trustworthy AI for Earth sciences requires codeveloping it with end users. For applications where AI can affect vulnerable or large populations, it's particularly important that AI developers working with end users employ a convergence approach—that is, have experts in the environmental, decision, and AI disciplines work together closely on specific, compelling problems.

AI2ES is developing and testing explainable AI methods to help describe to end users how AI models function. Existing physics-based prediction models have the advantage of being driven by the underlying physics of the problem; one can numerically represent the Navier–Stokes equations, for example. But because AI is unconstrained by the laws of physics, it could come up with a solution that violates those laws. Providing end users with different methods to understand what the AI model has learned may improve trust, and we are interviewing end users to understand the efficacy of those methods.

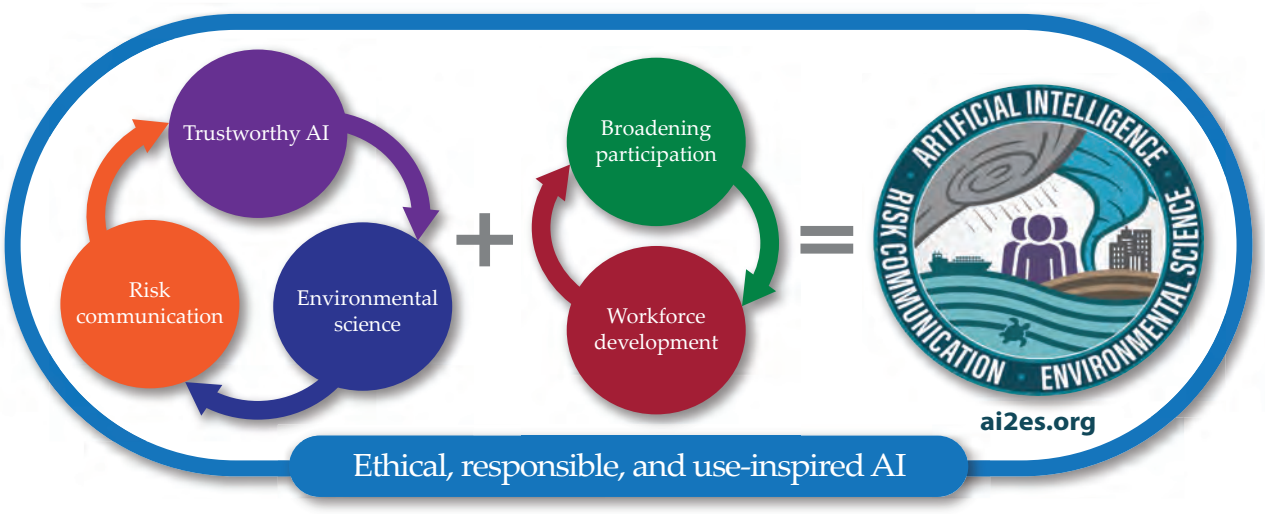
Trust, however, is contextual and subjective, and trust in AI models for weather and climate depends on a number of additional factors beyond peering inside the AI model. Those factors include having experience with the model over time, documenting performance and lack of bias across a range of extreme events for which the models are designed, and working with end users to ensure that their needs are met.

Saving sea turtles

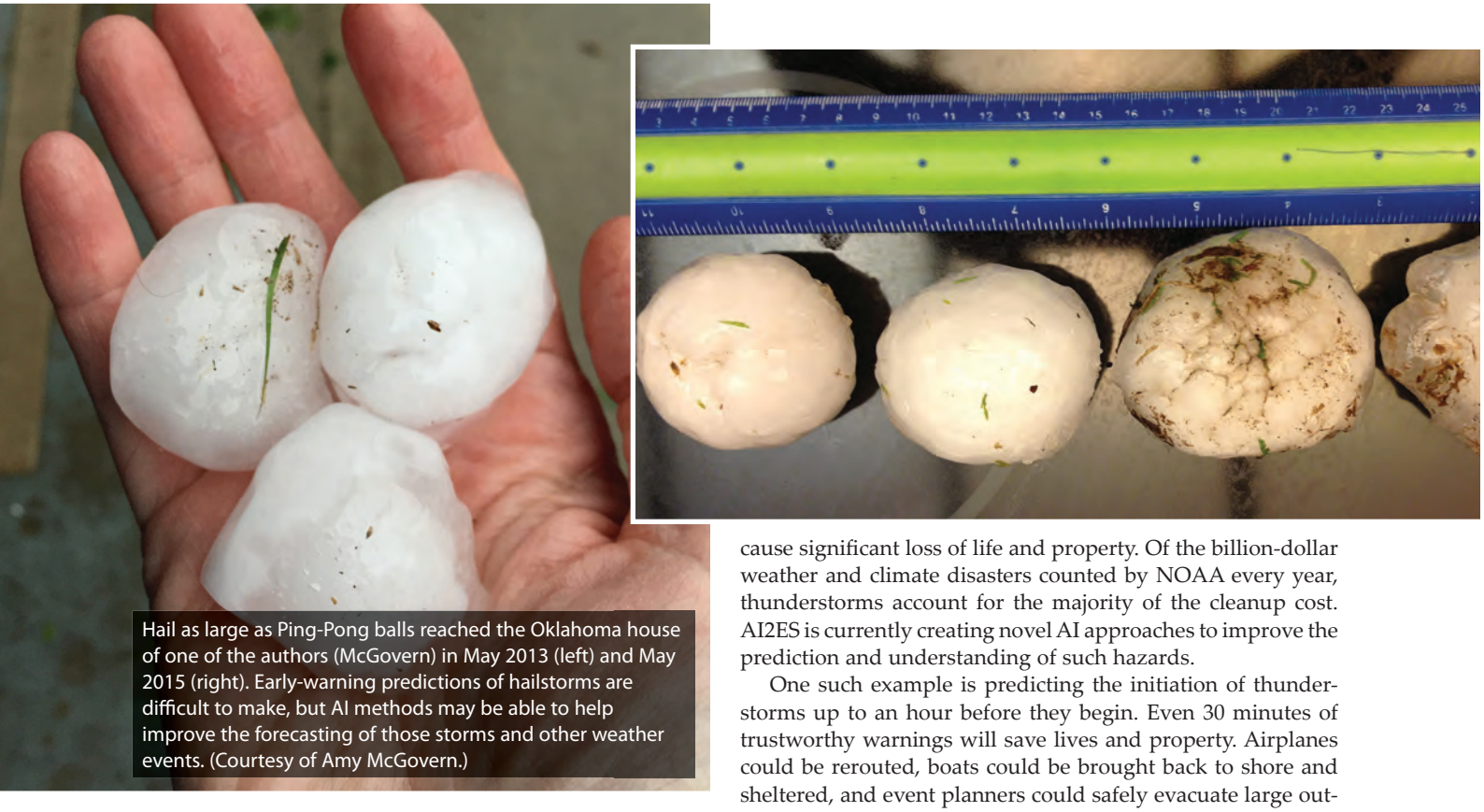
When strong cold fronts, such as the 2021 winter storm dubbed Uri, reach the southeast US, the temperatures of bays, lagunas, and other shallow bodies of water cool down rapidly. Below certain water temperature thresholds,[12] fish and endangered sea turtles become lethargic, or cold stunned, and most perish if they're not rescued. A community-wide effort for the Texas coast has grown since the mid 2000s to prepare for and mitigate the events. The program was updated following Uri, during which a record 13 000-plus sea turtles became cold stunned. Volunteers and employees of local, state, and federal agencies collect cold-stunned sea turtles along the shores or in bodies of water, and barge operators voluntarily interrupt their navigation through those waters. As climate change increases the frequency of extreme events, those types of large-scale organized human interventions will arguably need to become more frequent and more urgent if increasingly endangered species and fragile ecosystems are to be preserved.

To coordinate the rescue of cold-stunned turtles, a team needs real-time predictions of key environmental parameters, such as localized water temperature. When AI has access to time series of parameters from past extreme events, it is particularly well suited to develop targeted operational models, such as one for predicting when a cold-stunning event will happen. AI can take advantage of big, diverse data, such as gridded numerical weather predictions, satellite imagery, and ground-sensor readings.

Although the calibration of AI models can be lengthy, and care must be taken to maximize and test generalization, operational computations are fast once the information is available, particularly when done for just a few locations. The operational cold-stunning model is a type of neural network and has been used since the late 2000s. The first advisory and voluntary navigation interruption took place 8–10 January 2010 with a pre-



THE COMPREHENSIVE APPROACH created by AI2ES, the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography. (Courtesy of AI2ES.)



Hail as large as Ping-Pong balls reached the Oklahoma house of one of the authors (McGovern) in May 2013 (left) and May 2015 (right). Early-warning predictions of hailstorms are difficult to make, but AI methods may be able to help improve the forecasting of those storms and other weather events. (Courtesy of Amy McGovern.)

diction lead time of 48 hours. The system has been used several times since, including during the past three winters, with prediction lead times extended to 120 hours. The model is an essential decision tool that local, state, and federal agency representatives use when discussing with the private sector the optimal timing of activity interruptions in Texas's Laguna Madre. The specifically designed AI model provides the long lead time critical for redirecting cargo, contacting volunteers, and carrying out other actions.

The sea-turtle program brings the possibility to test how and why the trust in its AI model came about. The research team and end users are further developing AI ensemble models to quantify uncertainties around the predicted timing of the cold stunnings. An events' end is particularly challenging to predict with a longer lead time.

As the frequency of extreme events increases, sea levels rise, and other climate-driven challenges develop, even small flooding events will have large effects. So decision makers will have to start prioritizing and preparing for a broad range of emergency events beyond the largest ones, such as hurricanes, for which state and federal resources are deployed to assist local responders. Results are demonstrating that AI is a well-suited methodology to take advantage of large, diverse data sets and model the nonlinear processes of coastal zones and other environmental systems. Other coastal environmental models developed by AI2ES researchers include predictions of coastal fog,[13] coastal inundation, harmful algal blooms, eddy loop currents in the Gulf of Mexico, and compound flooding.

Severe storms

Thunderstorms worldwide produce various dangerous hazards: strong wind, lightning, hail, and tornadoes—all of which cause significant loss of life and property. Of the billion-dollar weather and climate disasters counted by NOAA every year, thunderstorms account for the majority of the cleanup cost. AI2ES is currently creating novel AI approaches to improve the prediction and understanding of such hazards.

One such example is predicting the initiation of thunderstorms up to an hour before they begin. Even 30 minutes of trustworthy warnings will save lives and property. Airplanes could be rerouted, boats could be brought back to shore and sheltered, and event planners could safely evacuate large outdoor events to avoid disasters, such as the hailstorm that hit Red Rocks Amphitheatre in Morrison, Colorado, in June and injured 80–90 people.

AI2ES's approach to modeling convective storms is codeveloped with researchers in NOAA's National Severe Storms Laboratory. Our work builds on NOAA's warn-on-forecast system (WoFS).[14] It is a numerical weather-prediction system that is run in real time at a high resolution over areas of the US where the Storm Prediction Center expects a higher probability of severe storms. AI2ES developed an AI postprocessing system that uses numerical weather-prediction models and current observations and outputs a real-time prediction of where storms are most likely to occur in the next 30 minutes. To help ensure that the system is trustworthy, AI2ES and NOAA will continue to develop it at NOAA's Hazardous Weather Testbed, a unique facility that allows forecasters and emergency managers to try out new technologies during severe weather events and to provide feedback to the developers.

AI2ES is also working to improve the understanding and prediction of tornadoes and hail. They are small-scale phenomena that are challenging to predict, especially on a short time scale and with high spatial precision, with current operational weather models. One of our most recent methods is codeveloped with NOAA researchers working on the WoFS. Our focus is on improving the nowcasting of severe hail events, which predicts such events at high resolution spatially and within an hour of their arrival. The WoFS runs in real time, but because of the computational complexity of the model, which ingests all the current observations, there is about a 15- to 30-minute lag between the observations and the system's predictions. We developed an AI prediction system that uses deep learning to combine WoFS predictions with data from the National Light-

ning Detection Network, operated by Vaisala,[15] and we demonstrated a significant improvement in the accuracy of short-term hail prediction.

Ethical, responsible AI

An integral part of trustworthy AI is ensuring that it is developed ethically and responsibly. If not, AI for environmental sciences can go wrong in numerous ways.[16] Extreme events tend to disproportionately harm areas with fewer resources and places with histories of systematic discrimination. It is critical that society ensures that AI is not deployed in any manner that will perpetuate environmental or climate injustices. That way, society as a whole can be more resilient to climate change.

Another potential issue with AI for weather prediction is bias, which affects all aspects of the AI training process. In recent work, we have developed a categorization of bias in AI for Earth sciences by breaking it into four main categories, each of which influences the others.[17]

- **Systemic and structural biases** include institutional and historical biases that can influence the choices of data that are made available, the labels on the data used for training AI, and other aspects of AI model development and use. For example, we demonstrated that tropical-cyclone initiation prediction is more likely to occur after sunrise than before because of institutional practices around examining the visible satellite imagery.
- **Data bias** can occur because of the data selected to train the models and the processing techniques used to prepare the data for training. Those choices can result in data that are not representative of the intended populations, areas, or events being modeled. Once the data are prepared and the AI model trained, biases can be present in the validation of the model. Humans must choose which score they will use to validate the model and which cases will be used as a case study. The choices can be affected by human judgment and decision biases, such as confirmation bias.[18]
- **Statistical and model biases** can affect the actual model that is trained and can be strongly affected by human biases. For example, human programmers must choose the methods that they will use to evaluate the model.
- **Human biases** are present throughout AI methods, from data selection to the choice of model, but they are also present in the deployment and use of the model. End users, such as forecasters and emergency managers, for example, may have information overload or may need to make split-second decisions, which can bias their use of AI.

Three of the perhaps most common ethical theories are applicable to AI for the environmental sciences: consequentialism, which judges the morality of an action by its consequences, such as through a benefit–cost analysis; deontology, which judges whether an act is ethical by how the act conforms to duties or moral principles, such as the imperative to be honest; and virtue ethics, which argues that a "right" action is important to achieve human well-being. Protecting the most vulnerable might not always pass a benefit–cost rule, but deontological and virtue ethics could require it, making it imperative.

But even to understand how AI models might affect specific decisions or users in particular circumstances generally requires an insider perspective, achievable only through developing AI with the people likely to be affected. Many of those concerns and needs can be addressed, and trustworthy AI can be developed by early and continued codevelopment of AI models with direct representation; meaningful, ongoing participation of likely end-user communities; and communication throughout the development process with risk-communication experts. But such capabilities require organizational intent from the teams developing the AI models.

The future of trustworthy AI

Given the current exponential growth of AI in the sciences, society stands at the cusp of major developments in AI for science and society in general. New methods could be developed and deployed with a swiftness that was not possible even a few years ago. That gives us an unprecedented opportunity to shape the process of how AI models are developed to fully benefit society and to address environmental and climate-justice issues. The process, however, must ensure that the models are ethical, responsible, and deserving of trust if society is to realize the full benefits of AI.

To achieve such goals, and to minimize problems during the release of new technology, more comprehensive processes and development teams must be engaged. Funding from federal agencies, private-sector entities, and other places must be structured to reflect those needs. Codevelopment of AI requires funding that allows for and encourages the development of multidisciplinary teams committed to working with end users. The benefits include acting ethically, avoiding large disparities, increasing resilience to climate change, and broadening the viewpoints, knowledge, and values represented on the modeling teams.

REFERENCES

1. S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson (2021).
2. A. McGovern et al., *Bull. Am. Meteorol. Soc.* **98**, 2073 (2017); S. E. Haupt et al., *Bull. Am. Meteorol. Soc.* **103**, E1351 (2022).
3. H. R. Glahn, D. A. Lowry, *J. Appl. Meteorol. Climatol.* **11**, 1203 (1972).
4. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press (2016).
5. R. J. Chase et al., *Weather Forecast.* **38**, 1271 (2023).
6. K. Bi et al., *Nature* **619**, 533 (2023).
7. A. McGovern et al., "Weathering environmental change through advances in AI," *Eos*, 28 July 2020.
8. R. E. Löfstedt, *Risk Management in Post-Trust Societies*, Palgrave Macmillan (2005); National Academies of Sciences, Engineering, and Medicine, *Communicating Science Effectively: A Research Agenda*, National Academies Press (2017).
9. National Academies of Sciences, Engineering, and Medicine, *Human-AI Teaming: State-of-the-Art and Research Needs*, National Academies Press (2022).
10. P. Slovic, *The Perception of Risk*, Routledge (2000).
11. E. Glikson, A. W. Woolley, *Acad. Manage. Ann.* **14**, 627 (2020).
12. D. J. Shaver et al., *PLoS One* **12**, e0173920 (2017).
13. H. Kamangir et al., *Mach. Learn. Appl.* **5**, 100038 (2021).
14. D. J. Stensrud et al., *Bull. Am. Meteorol. Soc.* **90**, 1487 (2009); K. A. Wilson et al., *Weather Clim. Soc.* **13**, 859 (2021).
15. H. Pohjola, A. Mäkelä, *Atmos. Res.* **123**, 117 (2013).
16. A. McGovern et al., *Environ. Data Sci.* **1**, e6 (2022).
17. A. McGovern et al., *Artif. Intell. Earth Sys.* **2**, e220077 (2023).
18. R. S. Nickerson, *Rev. Gen. Psychol.* **2**, 175 (1998).

PT