BAMS In Box

NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)

Amy McGovern, Ann Bostrom, Phillip Davis, Julie L. Demuth, Imme Ebert-Uphoff, Ruoying He, Jason Hickey, David John Gagne II, Nathan Snook, Jebb Q. Stewart, Christopher Thorncroft, Philippe Tissot, and John K. Williams

ABSTRACT: We introduce the National Science Foundation (NSF) AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES). This AI institute was funded in 2020 as part of a new initiative from the NSF to advance foundational AI research across a wide variety of domains. To date AI2ES is the only NSF AI institute focusing on environmental science applications. Our institute focuses on developing trustworthy AI methods for weather, climate, and coastal hazards. The AI methods will revolutionize our understanding and prediction of high-impact atmospheric and ocean science phenomena and will be utilized by diverse, professional user groups to reduce risks to society. In addition, we are creating novel educational paths, including a new degree program at a community college serving underrepresented minorities, to improve workforce diversity for both AI and environmental science.

KEYWORDS: Artificial intelligence; Atmosphere; Data science; Education; Ocean; Social Science

https://doi.org/10.1175/BAMS-D-21-0020.1

Corresponding author: Amy McGovern, amcgovern@ou.edu

In final form 22 February 2022

AFFILIATIONS: McGovern and Snook—University of Oklahoma, Norman, Oklahoma; **Bostrom**—University of Washington, Seattle, Washington; **Davis**—Del Mar College, Corpus Christi, Texas; **Demuth and Gagne**—National Center for Atmospheric Research, Boulder, Colorado; **Ebert-Uphoff**—Colorado State University, Fort Collins, Colorado; **He**—North Carolina State University, Raleigh, North Carolina; **Hickey**—Google, Mountain View, California; **Stewart**—National Oceanic and Atmospheric Administration, Boulder, Colorado; **Thorncroft**—University at Albany, State University of New York, Albany, New York; **Tissot**—Texas A&M–Corpus Christi, Corpus Christi, Texas; **Williams**—The Weather Company, an IBM Business, Andover, Massachusetts

While artificial intelligence (AI) has demonstrably improved prediction and understanding of many environmental science phenomena (e.g., Ahijevych et al. 2016; Williams et al. 2016; McGovern et al. 2017; Gagne et al. 2017, 2019; Lagerquist et al. 2019a; Barnes et al. 2019; Reichstein et al. 2019; Boukabara et al. 2021), there is often a lack of trust by environmental science decision-makers when it comes to relying on "black box" algorithms, especially in life-or-death situations (Karstens et al. 2018; Demuth et al. 2020). Developing AI that is trustworthy and useful for environmental risk management requires fundamental natural, mathematical, and social sciences research on the AI needs and perceptions of key users. These users' judgments and decisions may depend on their expertise and context (Larkin et al. 1980; Chi et al. 1981; Payne et al. 1992). Such research should include a users' understanding and perceptions of the AI method, its performance, and other factors emerging in empirical and theoretical research on AI (Mueller et al. 2019; Wang et al. 2019; Glikson and Woolley 2020).

We introduce the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), a national AI institute that conducts convergent research focused on creating trustworthy AI for the weather, climate, and ocean communities. AI2ES seeks to uniquely benefit humanity by developing novel physically based AI techniques that are demonstrated to be trustworthy, and to directly improve prediction, understanding, and communication of high-impact environmental hazards.

Developing trustworthy AI, particularly for the weather, water, and climate communities, is an *urgent and timely priority* (IPCC 2018; Reidmiller et al. 2018; ERISS Corporation and The Maritime Alliance 2019) at the highest levels of government and industry. NOAA has identified AI as a high priority in their strategic AI plan (NOAA 2019). Similarly, the White House continues to prioritize the development of innovative AI (National Science and Technology Council 2019b; Office of Science and Technology Policy 2019; National Science and Technology Council 2019a); the National Academies of Science, Engineering and Medicine cites improved forecasting of extreme events as a critical task (National Academies of Sciences, Engineering, and Medicine 2016); and NOAA and the National Weather Service (NWS) have increasingly focused on providing impact-based decision support services (IDSS) to reduce weather risks (Uccellini and Hoeve 2019), which requires developing new and improved forecast information to meet IDSS needs (Demuth et al. 2020). The European Centre for Medium-Range Weather Forecasts places high importance on improving predictions and understanding (European Centre for Medium-Range Weather Forecasts 2016). The recently released ECMWF 10-yr plan for AI states that "We anticipate that it will be increasingly difficult to distinguish between scientists working on machine learning and domain scientists in the future" (Düben et al. 2021).

AI algorithms for the weather, water, and climate communities must not only be skillful, but also trustworthy. The European Commission's report on Ethics Guidelines for Trustworthy AI notes that "AI systems need to be human-centric, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom" (High-Level Expert Group on AI 2019). AI2ES directly incorporates risk communication to connect the development of trustworthy AI to the human decision makers AI2ES aims to serve.

Before we proceed, a comment is in order regarding the terms machine learning (ML) versus AI. Machine learning is focused on the development of algorithms that allow computers to learn from data how to perform certain tasks without explicit programming. In contrast, artificial intelligence is a much broader concept that seeks to create algorithms providing human-like reasoning abilities. For the applications considered here we are only concerned with the machine learning type of algorithms. From here on we use both terms, ML and AI, interchangeably, always with the understanding that AI only refers to the machine learning style of AI.

Building trustworthy AI—Key components of AI2ES

Figure 1 provides an overview of the key components of AI2ES and the sidebar provides key terms and definitions. Our research cycle integrates development of foundational new AI/ML methods, working with atmospheric scientists and risk communication researchers in a virtuous cycle where each thread informs the others. Specifically we seek to develop trustworthy AI approaches for environmental science by

1) integrating risk communication research to determine which AI and explainable AI (XAI) features promote trustworthiness and use by different user groups, such as forecasters, for managing risk;

Fig. 1. The foundational research in trustworthy AI, environmental science, and risk communication forms a synergistic cycle where all parts interact with each other to inform the others. The blue circle around the diagram indicates our foundational focus on ensuring the AI is ethically and responsibly developed and applied as well as our key focus on use-inspired research. The broadening participation and workforce development components also synergize and together these comprise AI2ES. AI2ES welcomes new partners interested in working with us on any scale of the research, from foundational research to operations, or in broadening participation and workforce development. Interested researchers are welcome to join us for our AI2ES-wide presentations and to learn more about the different foci. In addition, we welcome additional private partnerships. To learn more, visit our website https://www.ai2es.org/.

2) leveraging and expanding approaches from physics-based AI and XAI;
3) raising awareness about negative side effects that AI has caused in other domains and helping the community to proactively avoid those;
4) using a variety of specific use cases (applications) to guide, test, and disseminate the AI approaches;
5) taking a multisector approach, involving academia, government agencies, and the private sector from the very beginning to facilitate maximal transition of the developed approaches from research to operations; and
6) creating new education pathways for AI, environmental science, and risk communication at all levels to improve workforce development and participation.

In the remainder of this article we give a quick overview of each of these components.

Convergent, virtuous cycle involving risk communication research with users

AI2ES is focused on developing trustworthy AI for professional users, such as weather forecasters, emergency managers, transportation officials, ecological and water resource managers. These professionals are close, direct users of AI forecast information, and their job responsibilities involve assessing risk and making decisions that have critical consequences for people's well-being. It is therefore important to study their AI interpretations, perceptions, uses, and needs, as situated in their varied decision-making contexts, in order to guide development and refinement of AI forecast information that is trustworthy.

The risk communication research in AI2ES brings to bear risk analysis, perception, and communication research theories and methods, with the goals of understanding how different features of AI and XAI influence trustworthiness of, trust in, and willingness to use AI guidance (e.g., Jacovi et al. 2021; Mueller et al. 2019; Glikson and Woolley 2020). This research is being conducted through structured interviews, experiments, and surveys in naturalistic settings for (i) different professional user groups, (ii) across different weather (severe convection, winter weather, tropical cyclones), coastal, and ocean hazards, and (iii) with different AI and XAI techniques. For example, in our initial, structured interviews, we are evaluating how weather forecasters' trust in AI guidance is influenced by the training approaches and datasets used for developing AI guidance, different bulk and case-study verification statistics, who developed the guidance, and interactivity with the output (including the ability to interrogate a source about how an answer was derived). Additional work will be conducted with XAI techniques to evaluate users' perceptions of, for example, algorithmic transparency, variable importance, and visualization techniques.

Conducting such user-oriented research is essential for multiple reasons. It will guide development of AI features and provision of AI guidance that aligns with users' key decision-making needs and contexts. Further, it will help improve users' evaluation and use of AI models and output, which intrinsically have limitations as all models do, to increase trust and use when warranted.

Leveraging physics-based AI and explainable AI

In addition to ensuring that the AI is skillful, AI2ES is developing new explainable and interpretable AI techniques (McGovern et al. 2019) focused on the needs of environmental science end users, developing new approaches to integrating the laws of physics into AI methods, developing novel ways to quantify and communicate model uncertainty, and ensuring that the AI is robust to natural and adversarial variations of AI model inputs.

As part of improving trust, many end users want to understand what the model has learned and if it is physically plausible (Jacovi et al. 2021). XAI provides an approach to looking inside the "black box" of ML models (Molnar 2018; McGovern et al. 2019), but many existing XAI

approaches do not adequately address environmental science phenomena such as spatiotemporal events. As noted above, our risk communication research agenda includes exploring professional users' perceptions and interpretations of XAI techniques applied to these and other similar contexts. Additional research on what constitutes good XAI from users' and human–AI teaming perspectives is needed (for examples and discussion of this see, e.g., Lu et al. 2020; Klein et al. 2021; Mueller et al. 2021; National Academies of Sciences, Engineering, and Medicine 2021; Schwalbe and Finzel 2021). We are developing new XAI approaches as well as investigating interpretable AI methods (Rudin 2019), which are designed from the start to be more human-understandable models. We are also ensuring that the AI and XAI approaches follow the laws of physics (e.g., Karniadakis et al. 2021; Jia et al. 2018, 2021), helping to further improve trust by ensuring the models cannot predict scenarios which are impossible and cannot learn nonphysical relationships for prediction (e.g., Yuval and O'Gorman 2020; Gettelman et al. 2021; Beucler et al. 2021a; Yuval et al. 2021). Murphy (1993) identifies three components of a good forecast: quality, or how well the forecast corresponds with observations; consistency, or how well the forecast corresponds with the forecaster's prior judgment; and value, or how much the forecast benefits users. Standard ML models focus on optimizing quality by fitting closely to observations often at the expense of consistency in space and time. Prior physical constraints on different parts of the ML pipeline have shown promise at improving consistency while maintaining or even further improving quality (e.g., Beucler et al. 2021b; Willard et al. 2020). While one could focus entirely on model skill, previous work has shown that forecasters and other scientific end users prefer a model based on physics.

Ethical and responsible AI and trust

"If not us, then who? If not now, then when? If not here, then where?" These inspiring words by Yeb Saño were spoken at the 2012 UN climate summit regarding the need to come together and combat climate change. We embrace these words to ensure that AI will be used in a responsible way in the environmental sciences. Namely, as the NSF center for creating *trustworthy* AI for environmental science, we are taking a lead on creating awareness and guidelines for the ethical and responsible use of AI for the weather, climate, and ocean (McGovern et al. 2022). Not only is it integral to the creation of trustworthy AI, but it is critical that we ensure that we are cognizant of and avoid unintentional negative consequences resulting from the introduction of AI. This in turn will avoid inadvertently creating—or increasing—environmental injustice through the use of AI.

Although it might seem as if the environmental sciences, due to their use of scientific methods and observations, are immune to the danger of increasing bias through AI, that is not the case. For example, while it might seem that the weather affects everyone equally, the effects of extreme weather are felt disproportionately by vulnerable communities and individuals (Environmental Protection Agency 2021). Likewise, if a tool uses data that are not available in areas where the predictions would be most needed, additional care needs to be taken to develop approaches that would work for all areas. It is well known that AI algorithms tend to reinforce and solidify unintentional biases in data (O'Neil 2016; Benjamin 2019). Given that we know there are existing unintentional biases in weather data, such as the population biases shown in hail and tornado reports (Allen and Tippett 2015; Potvin et al. 2019), one of the goals of AI2ES is to ensure that AI developers for weather, climate, and ocean applications have the knowledge and tools to create AI that can counteract these effects, to make the AI both ethical and responsible and to minimize bias. For example, we aim to develop a tool that would identify potential biases in data automatically to facilitate the developer counteracting these biases when training the AI model. This tool could identify that a dataset had a population bias or a bias toward specific sensors or specific times of data collection (all real examples of biases we have identified in weather and climate data) and encourage the

developer to counteract these biases through over/undersampling and data augmentation. Tools and principles such as these will help to ensure our AI is more trustworthy and will also help to address environmental justice needs (McGovern et al. 2022).

In the absence of empirical evidence of model interpretability by users, deontological ethics suggests that modelers have a duty to develop XAI that informs the user what the model is doing sufficiently to respect their decision-making autonomy. AI that convinces the user without actually informing them is manipulative rather than appropriately persuasive. This potential lack of honesty can create distrust (Lamb 2017). For example, generative adversarial networks (GANs; Goodfellow et al. 2014) and other generative AI models can create small-scale features in simulated imagery that look extremely realistic without necessarily being accurate to the same degree. This could potentially mislead forecasters, at least with regard to the appropriate level of confidence in the output images, and thus also in their interpretation of the model outputs. In contrast, XAI that informs user decisions thereby respects decision making autonomy, contributes to ethical AI, and should contribute to trustworthy AI. Five principles for science communication emerge from deontological ethics (Keohane et al. 2014), of which honesty is primary and imperative, the others being precision, audience relevance, process transparency, and specification of uncertainty about conclusions. All of these are relevant for XAI as seen as a form of science communication, and conform with findings to date regarding what might contribute to human trust in AI (e.g., Glikson and Woolley 2020).

Grounding AI development through use cases

AI2ES is principally focused on five environmental science applications: convective weather, winter weather, subseasonal to seasonal prediction, tropical cyclones, and coastal oceanography. We briefly outline our work in each of these areas. Note that the applications described here are not meant to cover all important areas in weather, climate, and coastal environments. Rather they should be seen as *case studies* that—while being important applications in their own right—serve the main purpose to *ground* the development of trustworthy AI methods in real-world environmental applications. Thus these topics were selected to cover many different types of problems (e.g., covering a large variety of meteorological phenomena, including a large range of temporal and spatial scales), requiring many different types of AI approaches [e.g., from generating simulated satellite imagery to using AI to learn new physics in subseasonal to seasonal prediction (S2S) applications], and to be supported by the expertise of the founding members (PIs and co-PIs). The variety of case studies and approaches offers further opportunity to ground the research and development with the aforementioned different professional users, yielding fundamental and applied research that is actionable.

Convective weather. In the area of convective weather, AI2ES is performing research on improving the skill and trustworthiness of AI predictions of weather hazards including tornadoes, hail, and severe wind. On the topic of tornado prediction (Lagerquist et al. 2019b), AI2ES is focusing on utilizing short-range (0–3 h) NWP forecasts and dual-polarization radar observations to both produce skillful tornado forecasts using deep learning and investigating how to best communicate these predictions to human forecasters. In the area of hail prediction (Gagne et al. 2017; Burke et al. 2020), AI2ES is investigating both short-range (0–6 h) and multiple-day (24–48 h) hail prediction. Hail prediction is also being used as a testbed for transition of AI2ES research results to industry applications.

The work in convective weather provides our initial test case for the synergistic research cycle involving AI, atmospheric scientists, and risk communication researchers. We are interviewing forecasters and emergency managers about their trust in several AI convective weather products and identifying how that trust varies as a function of XAI, visualization

and interactivity with the model, and model performance (Cains et al. 2022). The results of this will inform our AI model development for all of AI2ES.

Winter weather. Winter weather is a major hazard in the United States. Heavy snowfall, freezing rain, and extreme cold can all have severe impacts in many areas including travel (e.g., road, air, rail), utilities, commerce, and public health. AI2ES will be using AI to generate solutions to improve response and resilience to winter weather, with an early emphasis on road weather and related decision-making for public safety (e.g., through prediction of precipitation type and snow amounts) and environmental conservation (e.g., through improved efficiency of salt usage). We will also be addressing the needs of the National Weather Service and potentially other winter weather sensitive sectors including energy (e.g., utilities). AI-empowered winter weather analyses and predictions will be developed to provide trustworthy, customized weather information to support decision-making ahead of, during and after storm (recovery). The work will develop trustworthy products that exploit New York State Mesonet (NYSM; Brotzge et al. 2020), and the Oklahoma State Mesonet (McPherson et al. 2007), together with outputs from traditional observations and numerical weather prediction models as well as nontraditional user-provided data sources (e.g., road temperatures, snowplow speeds, salt activation, car sensors such as windshield wipers). AI is currently being used to extract weather information, such as visibility and precipitation, from the frequent camera images provided by the NYSM in a longer-term effort to extract such information from roadside cameras monitoring traffic and road conditions. The next phase of this work will exploit more data sources, to include more emphasis on precipitation type as they affect road conditions and decision-making needs. Automakers and insurers additionally are interested in this work as it improves safety and automation.

TCs. Proper representation of the convective structure of tropical cyclones (TCs) is important for the analysis and prediction of TC intensity and TC intensity change. However, existing satellites cannot observe TC convective structure at high temporal resolution. Namely, infrared imagery from geostationary satellites provides high spatial and temporal resolution, but upper-level cirrus obscures the underlying convective structure. In contrast, microwave imagery obtained from polar orbiting satellites reveals the TC convective structure, but has very low temporal resolution. Our AI2ES team seeks to use AI algorithms to combine the best of both worlds by learning to generate *simulated* microwave imagery from the geostationary imagery, thus *yielding imagery of TC convective structure at high temporal resolution*. In the next phase of the project we will seek to study this imagery to develop a better scientific understanding of the evolution of TC structure and to develop better prediction tools for TC intensity and intensity change (Slocum and Knaff 2020; Haynes et al. 2021). We also revisit the task of predicting TC intensification directly (without microwave imagery) using physics-based AI to address the challenge that TCs are behaving differently from year to year due to rising ocean temperatures, as documented by Schaffer et al. (2020).

S2S. Making predictions in the range of two weeks to two months is known to be particularly challenging (National Academies of Sciences, Engineering, and Medicine 2016). Furthermore, it has become clear that sometimes it is possible to have good forecast skill, but not at other times (Albers and Newman 2019; Mayer and Barnes 2021). Times at which good skill is possible provide *forecasts of opportunity* (Mariotti et al. 2020) and one of the goals of the AI2ES team is to use machine learning to identify those conditions under which good forecast skill exists, and then use XAI techniques to understand the physical processes at play. Research by members of the AI2ES team is exploring the concept of *abstention networks* to identify such conditions where skillful forecasts may be possible (Barnes and Barnes 2021a,b). Abstention

networks (Thulasidasan 2020) are neural networks that are trained to make predictions, but that additionally have the option to abstain, i.e., say "I do not know" in cases where their confidence for a skilled prediction is low. In doing so, the network is able to learn the more predictable behavior better than it would without abstention. Barnes and Barnes (2021a,b) explore how to apply abstention networks to Earth science applications, including S2S.

Coastal ocean environment. The coastal environment intersects oceans, land, and atmosphere and is home to critical ecosystems, large industrial facilities, and ports. Environmental datasets coming from in situ observing networks and satellite remote sensing platforms are still underutilized and represent a compelling opportunity to apply new AI methods to support better marine environment forecasting, science discovery, and stakeholder engagement. AI2ES is carrying out active research to combine physically based AI and machine learning with conventional numerical modeling to improve prediction skill and trustworthiness for a suite of coastal applications. These include problems ranging from marine ecology (e.g., predicting cold stuns to save sea turtles and fisheries), to marine transportation and offshore safety (e.g., marine fog forecast, ocean current and eddies prediction), to water quality (e.g., harmful algal blooms), to coastal hazards (e.g., compound flooding). And when coastal AI predictions significantly outperform existing models, such as for the prediction of coastal fog (Kamangir et al. 2021), XAI has the potential to bring new insights to the dynamics of the processes including air–sea interactions.

Taking a multisector approach

AI2ES takes a multisector approach (McGovern et al. 2020), one where researchers from academia, private industry, federally funded research centers, and government all work together to solve challenging problems. This approach represents the future for large-scale research initiatives as it brings together practitioners across the spectrum, from basic research all the way to operations. By working together, we can inspire new foundational research and transition research all the way to operations (R2O) as well as to other end users (R2X). For example, Google's flood forecasting work involves the cooperation of the Indian government as well as private industry and researchers (Matias 2021) and Schumacher et al. (2021) has demonstrated the critical need to work with directly with the targeted forecasters to develop an operational product.

As our climate changes, there are a number of wicked problems[1] that must be addressed using a convergent multisector paradigm (Bendito and Barrios 2016). For example, the changes to high-impact weather including severe storms, heat waves, drought, and torrential rain all require a collaborative approach to identifying the best long-term solutions that will facilitate climate resiliency and promote environmental justice. If only one agency or one sector studies the problems, they will not develop general solutions, and they may miss the inspiration of specific use cases to drive foundational research as well as opportunities to bring the research to end users through operational use.

[1] https://en.wikipedia.org/wiki/Wicked_problem

Workforce development and broadening participation

Broadening participation for both the AI and ES workforces is a major goal of AI2ES. We ensured that this goal was shared by all of our initial team members, including academic partners as well as private industry. We welcome additional partners and have already grown tremendously since our creation, including starting partnerships with NOAA Cooperative Science Centers and AMS's outreach and education programs.

AI2ES is developing and pilot testing an occupational skills award (OSA) for community colleges, a set of five classes open to a broad range of students, including nontraditional

ones, taking the students from basics to implementing AI projects within a geographic information system software application. This OSA award is being developed by Del Mar College with collaboration from Texas A&M University–Corpus Christi. Both are Hispanic serving and minority serving institutions and the goal of the OSA is to specifically create a pipeline of underrepresented minority students trained in AI and environmental science. The OSA debuted in fall 2021. Once the curriculum is tested, AI2ES will share it nationwide with other community colleges.

AI2ES is also developing multiple online workforce development modules. These include virtual summer schools, full university-level classes, and short courses. Our 2021 and 2022 summer schools focused on trustworthy AI for the environmental sciences. Our short courses are focused tutorial sessions, facilitating a deep dive into a specific topic. Each short course also includes example Jupyter notebooks so that participants can try out the ideas on their own environmental science phenomena. We recently completed a short course on XAI and have additional topics planned. Finally, university-level courses are being developed and shared online so that anyone around the world can use the material for learning and retraining. All of our material is available publicly on our website, ai2es.org.

Key terms

Below are our current working definitions of key terms. Note that as definitions in the literature vary and we gain additional experience, AI2ES is actively working to develop clear, shared definitions of these terms for use among our community.

- Explainable AI—An explainable AI method is one that can be explained post hoc, after training, in a way that makes it understandable (Schwalbe and Finzel 2021; Mueller et al. 2021). This includes methods to promote transparency into the black boxes, such as the ability to measure the importance of a variable or to see the effect of the values of that variable on the model as well as methods that allow a user to visualize patterns of activation in neural networks.
- Interpretable AI—An interpretable AI method is a model that is designed to be understood by humans without additional explanation. This does not include methods with large numbers of hyperparameters such as neural networks.
- Interactivity—The more interactive a method is, the more an end user can change parameters, select features, change weights on data points or parameters, visualize and select a specific model or ensemble of models, and change how they view the explanation and AI output (Rudin et al. 2022).
- Trustworthiness—Trustworthiness and trust are related, yet distinct, concepts. Trust is *relational*, in that it is "given to" or "placed in" someone or something, and trustworthiness is *evaluative*, in that it is a perceived characteristic of someone or something. With this in mind, trustworthiness is a (potential) trustor's evaluation, or perception, of whether, when, why, or to what degree someone or something should or should not be trusted. Current efforts to develop standards for trustworthiness (e.g., High-Level Expert Group on AI 2019) may lead some to confuse the broader concept of perceived trustworthiness with assessment of compliance with formal standards or policies for trustworthiness. A key distinction is that trustworthiness is a subjective evaluation that is largely dependent on the perceptions, values, experiences, and context of the assessor, which may or may not be influenced by standards or policies for trustworthiness.
- Deontological—Derived from the Greek word for duty (deon). Deontological ethics are rule-based ethics, or moral duties, such as the moral duty to be honest (Alexander and Moore 2021).

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant ICER-2019758.

References

Ahijevych, D., J. Pinto, J. Williams, and M. Steiner, 2016: Probabilistic forecasts of mesoscale convective system initiation using the random forest data mining technique. *Wea. Forecasting*, **31**, 581–599, https://doi.org/10.1175/WAF-D-15-0113.1.

Albers, J. R., and M. Newman, 2019: A priori identification of skillful extratropical subseasonal forecasts. *Geophys. Res. Lett.*, **46**, 12 527–12 536, https://doi.org/10.1029/2019GL085270.

Alexander, L., and M. Moore, 2021: Deontological ethics. Stanford University, https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/.

Allen, J., and M. Tippett, 2015: The characteristics of United States hail reports: 1955–2014. *Electron. J. Severe Storms Meteor.*, **10** (3), www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/149.

Barnes, E. A., and R. J. Barnes, 2021a: Controlled abstention neural networks for identifying skillful predictions for classification problems. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002573, https://doi.org/10.1029/2021MS002573.

——, and ——, 2021b: Controlled abstention neural networks for identifying skillful predictions for regression problems. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002575, https://doi.org/10.1029/2021MS002575.

——, J. W. Hurrell, I. Ebert-Uphoff, C. Anderson, and D. Anderson, 2019: Viewing forced climate patterns through an AI lens. *Geophys. Res. Lett.*, **46**, 13 389–13 398, https://doi.org/10.1029/2019GL084944.

Bendito, A., and E. Barrios, 2016: Convergent agency: Encouraging transdisciplinary approaches for effective climate change adaptation and disaster risk reduction. *Int. J. Disaster Risk Sci.*, **7**, 430–435, https://doi.org/10.1007/s13753-016-0102-9.

Benjamin, R., 2019: *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press, 172 pp.

Beucler, T., M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine, 2021a: Enforcing analytic constraints in neural networks emulating physical systems. *Phys. Rev. Lett.*, **126**, 098302, https://doi.org/10.1103/PhysRevLett.126.098302.

——, and Coauthors, 2021b: Climate-invariant machine learning. arXiv, 2112.08440, https://doi.org/10.48550/arXiv.2112.08440.

Boukabara, S.-A., and Coauthors, 2021: Outlook for exploiting artificial intelligence in the Earth and environmental sciences. *Bull. Amer. Meteor. Soc.*, **102**, E1016–E1032, https://doi.org/10.1175/BAMS-D-20-0031.1.

Brotzge, J. A., and Coauthors, 2020: A technical overview of the New York State Mesonet standard network. *J. Atmos. Oceanic Technol.*, **37**, 1827–1845, https://doi.org/10.1175/JTECH-D-19-0220.1.

Burke, A., N. Snook, D. Gagne, S. McCorkle, and A. McGovern, 2020: Calibration of machine learning–based probabilistic hail predictions for operational forecasting. *Wea. Forecasting*, **35**, 149–168, https://doi.org/10.1175/WAF-D-19-0105.1.

Cains, M. G., and Coauthors, 2022: NWS forecasters' perceptions and potential uses of trustworthy AI/ML for hazardous weather risks. *21st Conf. on Artificial Intelligence for Environmental Science*, Houston, TX, Amer. Meteor. Soc., 1.3., https://ams.confex.com/ams/102ANNUAL/meetingapp.cgi/Paper/393121.

Chi, M. T., P. J. Feltovich, and R. Glaser, 1981: Categorization and representation of physics problems by experts and novices. *Cognit. Sci.*, **5**, 121–152, https://doi.org/10.1207/s15516709cog0502_2.

Demuth, J., and Coauthors, 2020: Recommendations for developing useful and usable convection-allowing model ensemble information for NWS forecasters. *Wea. Forecasting*, **35**, 1381–1406, https://doi.org/10.1175/WAF-D-19-0108.1.

Düben, P., and Coauthors, 2021: Machine learning at ECMWF: A roadmap for the next 10 years. ECMWF Tech. Memo. 878, 20 pp., https://doi.org/10.21957/ge7ckgm.

Environmental Protection Agency, 2021: Climate change and social vulnerability in the United States: A focus on six impacts. EPA Tech. Rep. EPA 430-R-21-003, 101 pp., www.epa.gov/system/files/documents/2021-09/climate-vulnerability_september-2021_508.pdf.

ERISS Corporation and The Maritime Alliance, 2019: The ocean enterprise: A study of US business activity in ocean measurement, observation, and forecasting. NOAA Tech. Rep., 36 pp., https://cdn.ioos.noaa.gov/media/2017/12/oceanenterprise_feb2017_secure.pdf.

European Centre for Medium-Range Weather Forecasts, 2016: Strategy 2016–2025: The strength of a common goal. ECMWF Tech. Rep., 27 pp., www.ecmwf.int/sites/default/files/ECMWF_Strategy_2016-2025.pdf.

Gagne, D., A. McGovern, S. Haupt, R. Sobash, J. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, https://doi.org/10.1175/WAF-D-17-0010.1.

——, S. Haupt, D. Nychka, and G. Thompson, 2019: Interpretable deep learning for spatial analysis of severe hailstorms. *Mon. Wea. Rev.*, **147**, 2827–2845, https://doi.org/10.1175/MWR-D-18-0316.1.

Gettelman, A., D. J. Gagne, C.-C. Chen, M. W. Christensen, Z. J. Lebo, H. Morrison, and G. Gantos, 2021: Machine learning the warm rain process. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002268, https://doi.org/10.1029/2020MS002268.

Glikson, E., and A. W. Woolley, 2020: Human trust in artificial intelligence: Review of empirical research. *Acad. Manage. Ann.*, **14**, 627–660, https://doi.org/10.5465/annals.2018.0057.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, 2014: Generative adversarial nets. *28th Annual Conf. on Neural Information Processing Systems*, Montreal, QC, Canada, NeurIPS, 2672–2680, https://papers.nips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.

Haynes, K., C. Slocum, J. Knaff, K. Musgrave, and I. Ebert-Uphoff, 2021: Using machine learning to simulate 89-GHz imagery from geostationary satellites. *Third NOAA Workshop on Leveraging AI in Environmental Sciences*, Online, NOAA.

High-Level Expert Group on AI, 2019: Ethics guidelines for trustworthy AI. European Commission Rep., 41 pp., https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.

IPCC, 2018: *Global Warming of 1.5°C*. V. Masson-Delmotte et al., Eds., IPCC, 616 pp.

Jacovi, A., A. Marasović, T. Miller, and Y. Goldberg, 2021: Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. *Proc. 2021 ACM Conf. on Fairness, Accountability, and Transparency*, New York, NY, Association for Computing Machinery, 624–635, https://doi.org/10.1145/3442188.3445923.

Jia, X., A. Karpatne, J. Willard, M. Steinbach, J. Read, P. C. Hanson, H. A. Dugan, and V. Kumar, 2018: Physics guided recurrent neural networks for modeling dynamical systems: Application to monitoring water temperature and quality in lakes. arXiv, 1810.02880, https://doi.org/10.48550/arXiv.1810.02880.

——, J. Willard, A. Karpatne, J. S. Read, J. A. Zwart, M. Steinbach, and V. Kumar, 2021: Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles. *ACM/IMS Trans. Data Sci.*, **2**, 20, https://doi.org/10.1145/3447814.

Kamangir, H., W. Collins, P. Tissot, S. A. King, H. T. H. Dinh, N. Durham, and J. Rizzo, 2021: FogNet: A multiscale 3D CNN with double-branch dense block and attention mechanism for fog prediction. *Mach. Learn. Appl.*, **5**, 100038, https://doi.org/10.1016/j.mlwa.2021.100038.

Karniadakis, G., I. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, 2021: Physics-informed machine learning. *Nat. Rev. Phys.*, **3**, 422–440, https://doi.org/10.1038/s42254-021-00314-5.

Karstens, C., and Coauthors, 2018: Development of a human-machine mix for forecasting severe convective events. *Wea. Forecasting*, **33**, 715–737, https://doi.org/10.1175/WAF-D-17-0188.1.

Keohane, R. O., M. Lane, and M. Oppenheimer, 2014: The ethics of scientific communication under uncertainty. *Polit. Philos. Econ.*, **13**, 343–368, https://doi.org/10.1177/1470594X14538570.

Klein, G., R. Hoffman, and S. Mueller, 2021: Scorecard for self-explaining capabilities of AI systems. DARPA Tech. Rep., 9 pp.

Lagerquist, R., A. McGovern, and D. Gagne, 2019a: Deep learning for spatially explicit prediction of synoptic-scale fronts. *Wea. Forecasting*, **34**, 1137–1160, https://doi.org/10.1175/WAF-D-18-0183.1.

——, ——, C. Homeyer, D. Gagne, and T. Smith, 2019b: Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, **148**, 2837–2861, https://doi.org/10.1175/MWR-D-19-0372.1.

Lamb, M., 2017: *Ethics for Climate Change Communicators.* Oxford University Press, 42 pp.

Larkin, J., J. McDermott, D. P. Simon, and H. A. Simon, 1980: Expert and novice performance in solving physics problems. *Science*, **208**, 1335–1342, https://doi.org/10.1126/science.208.4450.1335.

Lu, J., D. D. Lee, T. W. Kim, and D. Danks, 2020: Good explanation for algorithmic transparency. *Proc. AAAI/ACM Conf. on AI, Ethics, and Society*, New York, NY, Association for Computing Machinery, 93, https://doi.org/10.1145/3375627.3375821.

Mariotti, A., and Coauthors, 2020: Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc.*, **101**, E608–E625, https://doi.org/10.1175/BAMS-D-18-0326.1.

Matias, Y., 2021: Expanding our ML-based flood forecasting. Google, https://blog.google/technology/ai/expanding-our-ml-based-flood-forecasting/.

Mayer, K. J., and E. A. Barnes, 2021: Subseasonal forecasts of opportunity identified by an explainable neural network. *Geophys. Res. Lett.*, **48**, e2020GL092092, https://doi.org/10.1029/2020GL092092.

McGovern, A., K. Elmore, D. Gagne, S. Haupt, C. Karstens, R. Lagerquist, T. Smith, and J. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, https://doi.org/10.1175/BAMS-D-16-0123.1.

——, R. Lagerquist, D. Gagne, G. Jergensen, K. Elmore, C. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, https://doi.org/10.1175/BAMS-D-18-0195.1.

——, and Coauthors, 2020: Weathering environmental change through advances in AI. *Eos*, **101**, https://doi.org/10.1029/2020EO147065.

——, I. Ebert-Uphoff, D. Gagne, and A. Bostrom, A. 2022: Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Environ. Data Sci.,* **1**, E6, https://doi.org/10.1017/eds.2022.5.

McPherson, R. A., and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **24**, 301–321, https://doi.org/10.1175/JTECH1976.1.

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

Mueller, S. T., R. R. Hoffman, W. Clancey, A. Emrey, and G. Klein, 2019: Explanation in human-AI systems: A literature meta-review synopsis of key ideas and publication and bibliography for explainable AI. arXiv, 1902.01876, https://doi.org/10.48550/arXiv.1902.01876.

——, E. S. Veinott, R. R. Hoffman, G. Klein, L. Alam, T. Mamun, and W. J. Clancey, 2021: Principles of explanation in human-AI systems. arXiv, 2102.04972, https://doi.org/10.48550/arXiv.2102.04972.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

National Academies of Sciences, Engineering, and Medicine, 2016: *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts.* National Academies Press, 350 pp.

——, 2021: *Human-AI Teaming: State of the Art and Research Needs.* National Academies Press, 140 pp.

National Science and Technology Council, 2019a: 2016–2019 progress report: Advancing artificial intelligence R&D. Executive Office of the President of the United States Rep., 48 pp., www.nitrd.gov/pubs/AI-Research-and-Development-Progress-Report-2016-2019.pdf.

——, 2019b: The National Artificial Intelligence Research and Development Strategic Plan: 2019 update. Executive Office of the President of the United States Rep., 50 pp., www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf.

NOAA, 2019: NOAA releases new strategies to apply emerging science and technology. NOAA, www.noaa.gov/media-release/noaa-releases-new-strategies-to-apply-emerging-science-and-technology.

Office of Science and Technology Policy, 2019: Accelerating America's leadership in artificial intelligence. White House, https://trumpwhitehouse.archives.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/.

O'Neil, C., 2016: *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, 272 pp.

Payne, J. W., J. R. Bettman, and E. J. Johnson, 1992: Behavioral decision research: A constructive processing perspective. *Annu. Rev. Psychol.*, **43**, 87–131, https://doi.org/10.1146/annurev.ps.43.020192.000511.

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

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, 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.

Reidmiller, D. R., C. Avery, D. Easterling, K. Kunkel, K. Lewis, T. Maycock, and B. Stewart, Eds., 2018: *Impacts, Risks, and Adaptation in the United States.* Fourth National Climate Assessment, Vol. II, U.S. Global Change Research Program, 1515 pp., https://doi.org/10.7930/NCA4.2018.

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

——, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, 2022: Interpretable machine learning: Fundamental principles and 10 grand challenges. *Stat. Surv.*, **16**, 1–85, https://doi.org/10.1214/21-SS133.

Schaffer, J. D., P. J. Roebber, and C. Evans, 2020: Development and evaluation of an evolutionary programming-based tropical cyclone intensity model. *Mon. Wea. Rev.*, **148**, 1951–1970, https://doi.org/10.1175/MWR-D-19-0346.1.

Schumacher, R. S., A. J. Hill, M. Klein, J. A. Nelson, M. J. Erickson, S. M. Trojniak, and G. R. Herman, 2021: From random forests to flood forecasts: A research to operations success story. *Bull. Amer. Meteor. Soc.*, **102**, E1742–E1755, https://doi.org/10.1175/BAMS-D-20-0186.1.

Schwalbe, G., and B. Finzel, 2021: XAI method properties: A (meta-)study. arXiv, 2105.07190, https://doi.org/10.48550/arXiv.2105.07190.

Slocum, C., and J. Knaff, 2020: Using geostationary imagery to peer through the clouds revealing hurricane structure. *100th AMS Annual Meeting*, Boston, MA, Amer. Meteor. Soc., J43.1, https://ams.confex.com/ams/2020Annual/webprogram/Paper369772.html.

Thulasidasan, S., 2020: Deep learning with abstention: Algorithms for robust training and predictive uncertainty. Ph.D. thesis, University of Washington, 186 pp.

Uccellini, L. W., and J. E. T. Hoeve, 2019: Evolving the national weather service to build a weather-ready nation: Connecting observations, forecasts, and warnings to decision makers through impact-based decision support services. *Bull. Amer. Meteor. Soc.*, **100**, 1923–1942, https://doi.org/10.1175/BAMS-D-18-0159.1.

Wang, D., Q. Yang, A. Abdul, and B. Y. Lim, 2019: Designing theory-driven user-centric explainable AI. *Proc. 2019 CHI Conf. on Human Factors in Computing Systems*, New York, NY, Association for Computing Machinery, https://doi.org/10.1145/3290605.3300831.

Willard, J., X. Jia, S. Xu, M. Steinbach, and V. Kumar, 2020: Integrating physics-based modeling with machine learning: A survey. arXiv, 2003.04919, https://doi.org/10.48550/arXiv.2003.04919.

Williams, J., P. Neilley, J. Koval, and J. McDonald, 2016: Adaptable regression method for ensemble consensus forecasting. *Proc. 30th AAAI Conf. on Artificial Intelligence*, Phoenix, AZ, AAAI, 3915–3921, www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12492.

Yuval, J., and P. O'Gorman, 2020: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nat. Commun.*, **11**, 3295, https://doi.org/10.1038/s41467-020-17142-3.

——, ——, and C. N. Hill, 2021: Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophys. Res. Lett.*, **48**, e2020GL091363, https://doi.org/10.1029/2020GL091363.