

<https://doi.org/10.1038/s44304-024-00014-x>

# The value of convergence research for developing trustworthy AI for weather, climate, and ocean hazards



Amy McGovern<sup>1,2,3,8</sup>✉, Julie Demuth<sup>3,4,8</sup>, Ann Bostrom<sup>3,5,8</sup>, Christopher D. Wirz<sup>3,4</sup>, Philippe E. Tissot<sup>3,6</sup>, Mariana G. Cains<sup>3,4</sup> & Kate D. Musgrave<sup>3,7</sup>

Artificial Intelligence applications are rapidly expanding across weather, climate, and natural hazards. AI can be used to assist with forecasting weather and climate risks, including forecasting both the chance that a hazard will occur and the negative impacts from it, which means AI can help protect lives, property, and livelihoods on a global scale in our changing climate. To ensure that we are achieving this goal, the AI must be developed to be trustworthy, which is a complex and multifaceted undertaking. We present our work from the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), where we are taking a convergence research approach. Our work deeply integrates across AI, environmental, and risk communication sciences. This involves collaboration with professional end-users to investigate how they assess the trustworthiness and usefulness of AI methods for forecasting natural hazards. In turn, we use this knowledge to develop AI that is more trustworthy. We discuss how and why end-users may trust or distrust AI methods for multiple natural hazards, including winter weather, tropical cyclones, severe storms, and coastal oceanography.

Of all the opportunities AI offers, reducing the climate change-enhanced risks of natural hazards is among those most urgent and within reach. Wildfires burning ever hotter and worsening compounding tropical cyclone and flooding events that disproportionately affect marginalized communities exemplify the increasing dangers and costs of natural hazard events<sup>1–3</sup>. At the same time, applications of AI to weather and climate have advanced rapidly [e.g., refs. 4–6]. Global data-driven weather models are making headlines for their speed, efficiency, and performance<sup>7–9</sup>, and it is clear there is great potential for AI to transform predictions for climate-related disasters. However, this potential brings a number of challenges.

Of all the challenges AI poses, trustworthiness is one of the biggest<sup>10,11</sup>. Trustworthiness is particularly critical for managing hazards that are being exacerbated by climate change and associated global governance failures<sup>12</sup>. Realizing the full potential of AI to help manage and mitigate increasing and

compounding risks without creating additional risks<sup>13,14</sup> requires a deep understanding of how to develop trustworthy AI.

Developing trustworthy AI is complex and requires a multifaceted approach that includes research that is (a) both fundamental and applied (i.e., in Pasteur's quadrant<sup>15</sup>), and (b) both disciplinary and convergent. Here, we illustrate how the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)<sup>16</sup> is conducting complementary research on these different fronts and how they work together to advance knowledge and practice about trustworthy AI. AI2ES is one of the 25 current NSF AI institutes. AI2ES researchers from AI, social, atmospheric, and ocean sciences are all working together to create trustworthy AI. To achieve this, we are working together to better understand what it means for AI to be trustworthy in the context of AI for weather and climate hazards (Fig. 1).

<sup>1</sup>School of Meteorology, University of Oklahoma, 120 David L Boren Blvd, Norman 73072 OK, USA. <sup>2</sup>School of Computer Science, University of Oklahoma, 110 W Boyd St, Norman 73019 OK, USA. <sup>3</sup>NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), Boulder, USA.

<sup>4</sup>Mesoscale & Microscale Meteorology Laboratory, NSF National Center for Atmospheric Research, 3090 Center Green Dr, Boulder 80301 CO, USA. <sup>5</sup>Daniel J. Evans School of Public Policy and Governance, University of Washington, 4105 George Washington Ln NE, Seattle 98195 WA, USA. <sup>6</sup>Conrad Blucher Institute, Texas A&M University-Corpus Christi, 6300 Ocean Drive, Corpus Christi 78412 TX, USA. <sup>7</sup>Cooperative Institute for Research in the Atmosphere, Colorado State University, 3925A West Laporte Ave., Fort Collins 80521 CO, USA. <sup>8</sup>These authors contributed equally: Amy McGovern, Julie Demuth, and Ann Bostrom.

✉e-mail: [amcgovern@ou.edu](mailto:amcgovern@ou.edu)



**Fig. 1** | Convergence research as instantiated in AI2ES.

## Convergence research for trustworthy AI

Over the last decade, the value of convergence research has emerged in recognition of increasingly complex and societally important science problems. These problems require knowledge creation and innovation beyond the bounds of traditional science disciplines and thus demonstrate the need for research that deeply integrates intellectually diverse sciences<sup>17–20</sup>. The need for and value of convergence research has been documented in the context of natural hazards, disasters, and risks<sup>21–23</sup>. Furthermore, there have been broad calls for facilitating convergence of natural, social, and computational sciences to advance Earth systems science<sup>24</sup>.

Developing AI for weather, climate, and coastal hazards that is trustworthy, as well as trusted and used, is a multifaceted science problem that requires convergence research to meaningfully and comprehensively address. To the best of our current understanding, trustworthiness stems from a complex intersection of factors, including users' decision-making needs and contexts; data quality and representativeness; model development processes, techniques, and specifics; model availability, interpretability, explainability (tools to aid comprehension), and integration into users' workflows; perceptions of the model developers' expertise; and model skill across hazards and geography<sup>25–28</sup>. The centrality of human perceptions, communications, and decisions at this intersection makes the inclusion of social, behavioral, and cognitive scientists critical. Accordingly, AI2ES brings together geoscientists with domain expertise; computational scientists with diverse AI expertise; and social/behavioral scientists with expertise in judgment and decision making, risk and decision analysis, and communication of natural hazards. We are collectively committed to partnering with professionals like weather forecasters, whose work affects society broadly and is increasingly likely to depend on AI. Moreover, several AI2ES scientists are knowledge brokers and boundary spanners, who bridge across fields as well as between research and practice<sup>29,30</sup>. Together these attributes and commitments facilitate generative, supportive, and innovative collaborations amongst the team. Figure 1 illustrates how we work together across hazards and researchers to conduct convergence research.

In this paper we illustrate our approaches to trustworthy AI through three research threads. These three examples each highlight different research components and problem-solving strategies. They are representative of the efforts in AI2ES, but we have many additional examples as well.

## AI trustworthiness perceptions of professional decision-makers

Risk messages consist of any type of information about the potential threat, severity, consequences, and recommended actions that pertain to a risk, including those posed by natural hazards<sup>31</sup>. In the context of weather, climate, and coastal hazards, the predictive information provided by AI models about whether, when, and where hazards will occur, along with their potential magnitude and impacts, therefore constitutes a type of risk information.

AI as a form of risk information serves a particular role for professional, public, and private sector decision-makers. These include weather forecasters and broadcasters, emergency managers, water resource managers, critical infrastructure officials, school officials, and others. These decision-makers hold the responsibility to protect the lives and livelihoods of their staff, their constituents, and society. Such professional users access, interpret, and use AI-derived information to assess natural hazard risks and to make job role-specific decisions, such as threat, communication, mitigation and protective actions, and emergency response. For example, in the event of hazardous weather conditions, critical infrastructure officials may be responsible for shutting down air or marine traffic, closing roads, or handling power outages. Emergency managers can be responsible for communicating warnings, informing or implementing protective actions such as evacuations or sheltering in place, and correcting misinformation.

AI trustworthiness and trust are paramount for such professional users because of the high-stakes nature of their job decisions, which often are further characterized by high uncertainty and time pressures. Such decision contexts render professional users vulnerable to the AI information that guides their decision-making. This amplifies the criticality of their ability to assess whether they can rely on and put faith in that AI information. These factors manifest acutely in the context of severe convective weather (including tornadoes, high winds, and hail), which thus serve as an important, valuable use case for initial research on AI trustworthiness and trust.

In our first example, we leveraged two newly developed, prototype AI models for severe convective weather prediction. The first applies a random forest-based technique to predict the probability of 1" and 2" hail<sup>32</sup>. The second is a 2-dimensional convolutional neural net trained to predict the probability of convective storm mode being supercellular, quasi-linear, or disorganized<sup>33</sup>. These existing prototypes allowed our working group to design formative research and collect data with National Weather Service forecasters to explore foundational research questions about AI trustworthiness with concrete examples. For instance, a central research question that we have investigated is how different descriptive and performance attributes of new AI guidance influence forecasters' perceptions of a model's trustworthiness. These include the AI technique used, the training of the AI model, the AI model input variables, the performance of the AI model, and the interactivity with the AI model. As a group of risk communication scientists collaborating closely with atmospheric and AI scientists, we co-developed a survey and structured interview protocol with a decision task. Using the survey and interview protocol, we examined forecasters' assessments of how essential different features are when gaining familiarity with and when operationally using new forecast guidance; how forecasters perceived and evaluated the information provided to them about the prototype attributes; and how they applied such information to assess the trustworthiness of AI-based forecasting guidance.

Analysis of our data yielded multiple findings at different knowledge scales about trustworthiness and trust of AI information for high-stakes decision-making (see<sup>27</sup> for full details). For instance, at the prototype-specific scale, we found that developers hand-labeling the inputs for the AI storm mode guidance increased forecasters' stated trustworthiness of it. However, this was contingent on the labelers having the relevant domain expertise to do so, because storm mode is a complicated, latent feature to characterize and identify. This finding suggests that the resource-intensive task of human hand-labeling for developing predictive AI guidance may be important for user trustworthiness for difficult but important tasks. At a broader scale, across both prototype products and attributes we evaluated, we found that information about the AI model technique (particularly about

the model input variables), information about the model performance (particularly about failure modes), and being able to interact with the AI model output were all essential to increasing forecasters' assessment of the trustworthiness of the AI guidance. Finally, at the most generalized scale, we found that forecasters' trust in new AI guidance is a progressive process. This process includes key phases of their initial exposure to the new guidance, their continued familiarization with it through non-operational exploration, and their experience in the operational forecasting environment when they can observe the AI model performance in real-time and potentially use it for forecasting. We further found that centering the forecaster by including them in the research process is a critical part of trustworthiness-building, and we recommended mechanisms for furthering this work.

Our complementary research includes a conceptual framework that recognizes trust and trustworthiness as relational concepts that require centering the one who is doing the trusting<sup>34</sup> and a research agenda that emphasizes user-centered perspectives on AI in environmental sciences<sup>36</sup>.

These findings illustrate the value of our convergence science approach and the multitude of results, across different scales, that resulted from it.

### New ways of thinking about science

The AI2ES working group studying severe convective weather prediction is one of several use-case-inspired working groups in AI2ES. Like many of our working groups, the group studying severe convective weather prediction integrates the work of foundational AI experts and geoscientists to develop user-oriented solutions to the challenge of trustworthy AI. This is done in collaboration with, and as identified and defined by, risk communication scientists. Diverse teams such as these are well situated for new ways of thinking about their scientific problems: they have the expertise and diverse perspectives to apply close analogies and create new, shared understandings of unexpected outcomes and surprises<sup>35</sup>. Being willing to engage in such team work is an essential characteristic of convergence science. Being clear about research goals and being willing to change goals and look for new hypotheses in response to inconsistent findings can facilitate scientific breakthroughs<sup>35,36</sup>.

Convergence research efforts in AI2ES to date have fostered several new research directions and ideas. Working groups with representation from across the disciplines and organizations in AI2ES have focused on specific environmental hazard use cases (e.g. the prediction of coastal fog<sup>37</sup>) to co-develop research methods, instruments, and products that will meaningfully advance both disciplinary sciences and convergence science. For example, identifying different comparatives to verify performance for ocean loop current eddy AI/ML models was a challenge for the oceanography research group, which was grappling with the lack of in situ observational data in this domain. In this domain, as in climate modeling, it is common to compare the performance of predictive models with reanalyses, rather than directly with environmental observations. After discussions about the role of verification in trustworthiness, the team ultimately got permission to use drifter data from another research group as a verification comparative, in addition to comparisons against reanalysis.

Convergent problem-solving has also inspired renewed consideration of probabilistic as well as deterministic approaches for coastal fog predictions and has expanded perspectives on the value of physics-informed explainable (X)AI models<sup>38</sup>. Challenges that we have tackled include how to create new architectures to extract spatio-temporal patterns involving multiple physical variables and developing XAI methods. XAI can be used to better understand and quantify the roots of the improved performance as well as to better explain the mechanics of the model to end-users.

Team problem-solving has also led to co-development of more rigorous research methods by borrowing ideas across the different fields. For example, one working group co-developed data analysis procedures by borrowing rigorous social science content analysis methods to improve the reliability and validity of hand labeling for supervised ML modeling of road surface conditions<sup>39</sup>. As another example, based on feedback from a large (100+) user base, AI2ES is developing uncertainty quantification methods and

visualizations to help predict and communicate the start and end of cold stunning events during which endangered and threatened sea turtles and other marine organisms become lethargic and must be rescued to survive<sup>6</sup>. Foundational AI scientists contributed adaptation of probabilistic neural networks while geoscientists explored the uncertainty related to the training of the models and its predictors, and risk communication scientists formalized and are systematically studying users' interactions with these models.

All of these examples exemplify innovative ways of thinking and new ideas that have emerged at the intersection of the environmental, computational, data, and risk communication sciences in AI2ES around shared research problems and tasks.

### Using AI for foundational science discovery

Creating trustworthy AI requires us to develop and evaluate a wide variety of AI techniques, including explainable and interpretable AI [e.g., ref. 40] and physics-based AI [e.g., refs. 41, 42]. By developing and testing these on a wide variety of applications, we have a suite of techniques that we can then use for novel applications, including foundational science discovery. Human-AI teams can make use of AI techniques for science discovery in a variety of ways. We discuss several ways that AI2ES is using AI for foundational knowledge discovery.

Many of the most destructive weather and climate hazards are both rare in terms of frequency and also rare in terms of rich datasets describing these hazards. One such example is that of tropical cyclones, and in particular data on rapid intensification. By definition, rapid intensification is a rare event since it requires exceeding the 90th percentile of tropical cyclone intensification within 24 h. Complicating the problem of rarity is the lack of observational data of the internal structure of the tropical cyclones. In order to observe the internal structure of the tropical cyclone, you either need a fly-through by aircraft reconnaissance (rare and only provides a limited view of the overall storm structure over a few hours)<sup>43</sup> or to have observations from an orbiting microwave sensor that can peer inside the clouds and observe the full structure of the storm. However, the microwave sensors are on low earth orbiting satellites instead of geostationary, so their passes are quick and infrequent<sup>44,45</sup>.

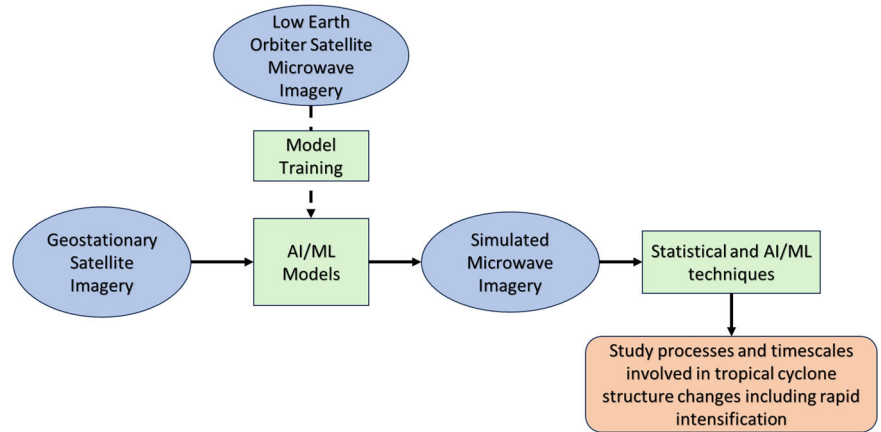
We have created a deep learning microwave sensor that transforms geostationary satellite imagery into a simulated microwave sensor, specifically for the 89-GHz channel<sup>46</sup>. This approach (Fig. 2) allows us to take advantage of the temporal and spatial resolution of the current suite of geostationary satellites to study the evolution of the internal structure of tropical cyclones through simulated microwave imagery. In particular the simulated microwave sensor allows for the study of processes involved in rapid changes of tropical cyclone structure and intensity, which have traditionally been missed by infrequent overpasses of low earth orbiting satellites with microwave sensors. While caution is needed in the interpretation of results based on uncertainties in the simulated microwave imagery, the generated archive of tropical cyclone overpasses is allowing for studies using both traditional statistical and AI/ML methods into these rare and data-sparse events. This new data that will also facilitate discovery of the coupled atmospheric-ocean processes that lead to rapid intensification.

Another example that we have focused on in AI2ES are tornadoes. Tornadoes are rare and they are also under-sensed in that existing instruments such as radar physically cannot provide a full 3-dimensional picture of the atmosphere. While studying such observational data is a critical step in understanding tornadogenesis, or why one storm generates a tornado when a similar storm does not, understanding the full behavior of the atmosphere requires creating idealized simulations. These simulations provide a complete picture of the atmosphere but they have far too much data for a human to analyze. AI techniques enable us to look for patterns in large-scale data sets, such as idealized simulations of tornadoes.

### Convergence practices in AI2ES

Convergence research is an ongoing, evolving process that functions differently during different phases of a research team's tenure and research efforts. Sundstrom et al.<sup>19</sup> characterize this process as cycling between

**Fig. 2 | Use of simulated microwave dataset to study tropical cyclone structure.** AI/ML models are trained with microwave imagery from low earth orbiting satellites as truth to use geostationary satellite input to generate simulated microwave imagery. This simulated microwave imagery is then used as a dataset to study tropical cyclone structure and rapid intensification.



transcendent and focused phases of research, in which the former involves expansive ideating about important research problems to address and the latter involves more refined approaches to pursue specific research threads. This cyclic approach that tacks between larger- and finer-scale science while also building on an evolving foundation of research efforts is particularly valuable in the AI domain, in which the landscape of AI awareness, possibilities, and progress is rapidly evolving. Indeed, convergence research approaches are considered to be particularly well-suited to nonstationary systems<sup>19</sup>, as the complex, nonlinear aspects of AI for Earth systems science represents.

Equally important to the pressing, innovative science problems AI2ES is addressing, however, are the practices and values by which we work to facilitate convergence. Convergence requires leadership that convenes and enables full intellectual and equally valued participation from diverse and often disparate disciplines. It requires significant investments of time: time for relationship-building among team members; time for exchanging, translating, blending, and connecting ideas; and time for synthesizing co-produced knowledge. Convergence research also requires patience and persistence, given the multitude of challenges that inevitably arise. These and other requirements have been documented by scholars such as Peek et al.<sup>21</sup>, Morss et al.<sup>23</sup>, and Sundstrom et al.<sup>19</sup>.

In AI2ES, we operationalize these essential practices and values in a number of ways. We hold biweekly meetings of our leadership team of 15 people comprised of the Institute PI, co-PIs, focus area leads, private and government sector lead collaborators, and our external evaluator. In these leadership meetings, a culture is fostered that values convergence research, each other, and all AI2ES personnel. We also hold biweekly site-wide meetings to reach all of AI2ES and external colleagues, to foster inclusion and broad reach and to share the diversity of science that is emerging from and relevant to AI2ES. At the “working level”<sup>23</sup>, we have multiple, cross-cutting working groups that are organized around key research and workflow topics of natural hazard use cases. These working groups functionally facilitate transcendent and focused phases of convergence research and practices, in which we collaboratively and iteratively generate and refine research questions, develop and execute research steps, interpret results, and create outputs for dissemination.

Supplementing organized and planned working groups, AI2ES also fosters convergence through emergent working groups and networks among its early career researchers (undergrads, grad students, postdocs, and new research scientists and faculty). These small interdisciplinary and inter-institutional working groups develop initially as mutual support for specific research tasks, but often transition into fruitful environments to ask questions, share experiences, give advice, and transfer knowledge across typically strong disciplinary and institutional divides. These nurturing environments are key for our convergence science, and they are essential for developing the skills and interdisciplinary networks of the next generation of the AI workforce.

We also facilitate convergence through the collaborative creation of a variety of *boundary objects*<sup>47</sup>. Examples of boundary objects for AI2ES include a co-developed figure illustrating the types of biases likely to affect AI development for Earth Sciences<sup>13,14</sup>; a co-developed codebook for rigorous quantitative content analysis of images for hand-labeling precipitation by an interdisciplinary group of scientists as input into an AI model<sup>39</sup>; and a co-developed structured interview protocol that distills and conveys the main AI model attributes to be evaluated<sup>27</sup>.

Importantly, AI2ES’s convergence research efforts also are built on a foundation of prior experience from many on the AI2ES team who have participated in and/or led interdisciplinary and convergence research. This reveals another contribution of AI2ES and reflects a broader and increasing need pertaining to natural hazards, AI, and society – that is, of training both the current and future workforce about the needs for, approaches for, and utility of deeply working in highly disciplinarily diverse teams.

Another critical piece of convergence research is education. If we are to take advantage of and deploy more convergent efforts to tackle our timely and wicked challenges<sup>48</sup>, the training of our current and future scientists must evolve beyond current practices<sup>49,50</sup>. Current practices at most universities focus on disciplinary knowledge, often to the exclusion of interdisciplinary and convergence science. With the goal of training future convergence researchers, AI2ES trains students, at the undergraduate and graduate level, to work in interdisciplinary research from the start. In order to achieve this, we intentionally create an inclusive culture where each member of the research group knows they are able to ask questions of the other disciplines and are willing to learn. Assembling and nurturing diverse, convergence research teams that involve mentors and postdoctoral or graduate scholars presently requires that the team be relatively large, preferably involving at least some established collaborations. AI2ES is developing a better understanding of which components are essential to foster convergence research teams and nurture future scientists at the intersection of AI, geosciences, and risk communication.

AI2ES also trains the existing workforce. This includes cross-training existing professionals in new skills such as AI but it also includes training those professionals labeled earlier as knowledge brokers or boundary spanners. In AI2ES, we support scientists across multiple career stages toward spanning boundaries while deepening their disciplinary expertise. Learning to span boundaries requires taking a leap of faith for many scientists, which is easier to do when they are in a supportive environment where they can see the success of other such professionals.

## Future work

The examples discussed above are illustrative, although not fully encompassing, of the convergence research that AI2ES is conducting and producing. Furthermore, AI2ES efforts are ongoing and, arguably, just beginning.

Looking to the future of trustworthy AI for weather and climate, one of the key new developments in AI recently for weather has been the



emergence of global data-driven weather models<sup>7–9</sup>. AI2ES is collaborating with the Cooperative Institute for Research in the Atmosphere at Colorado State University and the NOAA Global Systems Lab on developing research methods to evaluate the performance of these models. Our approach derives from AI2ES convergence research and other human–AI teaming research in the weather and climate domain. The goal is to understand how weather forecasters and other critical users perceive the trustworthiness of purely data-driven models relative to other AI/ML models, as well as to more conventional numerical weather prediction models. For example, taking the forecaster’s perspective entails focusing on evaluating model performance on metrics that are more specific to the meteorological functions and decision-support purposes of such models as opposed to generic metrics such as root mean square error.

Another key part of trustworthy AI is ensuring that the AI is developed and deployed ethically and responsibly. AI2ES is leading an effort on ethical and responsible AI for Earth and environmental sciences. We have first highlighted many of the issues that can arise from improper development and use of AI<sup>13</sup> and have recently created a classification system focused on bias in both data and AI models McGovern et al.<sup>14</sup>. Our current efforts on ethical and responsible AI are centered on demonstrating how to mitigate the various types of bias that we highlighted.

Moving forward, partnerships will be key to ensuring that AI is trustworthy for the full cycle of development, from the foundational research to deployment. Currently private industry is leading the development of the foundation models while government and public entities lead the collection of data<sup>51</sup>. Academia and government may not have sufficient computational resources to develop large foundation models and keep pace with market forces and the private sector. Very rapid development and proprietary methods in the private sector can complicate partnerships and make it difficult to evaluate AI models, let alone regulate them. Academia has the ability take larger risks in initial research, and government can help engage critical end-users in model co-production. Yet such partnerships can be challenging to develop and sustain, especially in the United States where funding has typically not been set up to facilitate connections and merge efforts across academia, private industry, and governmental organizations. The challenges but importance of these partnerships pose another important opportunity for future convergence research efforts.

Received: 9 February 2024; Accepted: 20 April 2024;

Published online: 01 July 2024

## References

1. Suran, M. Raging wildfires are exposing more people to smoky air—here’s what that means for health. *JAMA* **330**, 1311–1314 (2023).
2. Simpson, N. P. et al. Adaptation to compound climate risks: a systematic global stocktake. *iScience* **26**, 105926 (2023).
3. Singh, D. et al. Focus on compound events. In *Fifth National Climate Assessment* (U.S. Global Change Research Program, Washington, DC, USA, 2023).
4. Lopez-Gomez, I., McGovern, A., Agrawal, S. & Hickey, J. Global extreme heat forecasting using neural weather models. *Artif. Int. Earth Syst.* **2**, e220035 (2023).
5. Chattopadhyay, A., Gray, M., Wu, T., Lowe, A. B. & He, R. Oceannet: A principled neural operator-based digital twin for regional oceans. *arXiv* <https://doi.org/10.48550/arXiv.2310.00813> (2023).
6. McGovern, A., Tissot, P. & Bostrom, A. Developing trustworthy AI for weather and climate. *Phys. Today* **77**, 26–31 (2024).
7. Ebert-Uphoff, I. & Hilburn, K. The outlook for AI weather prediction. *Nature* <https://doi.org/10.1038/d41586-023-02084-9> (2023).
8. Bi, K. et al. Accurate medium-range global weather forecasting with 3d neural networks. *Nature* **619**, 533–538 (2023).
9. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).
10. Future of Life Institute. *Pause Giant AI Experiments: An Open Letter*. <https://futureoflife.org/open-letter/pause-giantai-experiments/> (2023).
11. White House, T. *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. <https://www.federalregister.gov/d/2023-24283> (2023).
12. Berrang-Ford, L. et al. A systematic global stocktake of evidence on human adaptation to climate change. *Nat. Clim. Change* **11**, 989–1000 (2021).
13. McGovern, A., Ebert-Uphoff, I., Gagne, D. J. & Bostrom, A. Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Env. Data Sci.* **1**, e6 (2022).
14. McGovern, A. et al. Identifying and categorizing bias in AI/ML for earth sciences. *B. Am. Meteorol. Soc.* **105**, E567–E583 (2024).
15. Stokes, D. E. *Pasteur’s Quadrant: Basic Science and Technological Innovation* 196 (Brookings Institution Press, Washington DC, 1997).
16. McGovern, A. et al. NSF AI Institute for Research on trustworthy AI in weather, climate, and coastal oceanography (AI2ES). *B. Am. Meteorol. Soc.* **103**, E1658–E1668 (2022).
17. Roco, M., Bainbridge, W., Tonn, B. & Whitesides, G. *Converging Knowledge, Technology, and Society: Beyond Convergence of Nano-Bio-Info-Cognitive Technologies* 450 (Dordrecht, Heidelberg, New York, London, 2013).
18. National Academies of Sciences Engineering and Medicine. *Fostering the Culture of Convergence in Research: Proceedings of a Workshop* 82 (The National Academies Press, Washington, DC, 2019).
19. Sundstrom, S. M. et al. The emergence of convergence. *Sci. Anthropol.* **11**, 00128 (2022).
20. National Science Foundation. *Learn About Convergence Research*. <https://new.nsf.gov/funding/learn/research-types/learn-about-convergence-research> (2024).
21. Peek, L., Tobin, J., Adams, R. M., Wu, H. & Mathews, M. C. A framework for convergence research in the hazards and disaster field: the natural hazards engineering research infrastructure converge facility. *Front. Built Environ.* **6**, 110 (2020).
22. Lakhina, S. J., Sutley, E. J. & Wilson, J. “how do we actually do convergence” for disaster resilience? cases from Australia and the United states. *Int. J. Disaster Risk Sci.* **12**, 299–311 (2021).
23. Morss, R. E., Lazrus, H. & Demuth, J. L. The “inter” within interdisciplinary research: strategies for building integration across fields. *Risk Analysis* **41**, 1152–1161 (2021).
24. National Academies of Sciences Engineering and Medicine. *Next generation Earth systems science at the National Science Foundation* 136 (The National Academies Press, Washington DC, 2021).
25. Jacovi, A., Marasović, A., Miller, T. & Goldberg, Y. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 624–635 (ACM, 2021).
26. Bostrom, A. et al. Trust and trustworthy artificial intelligence: a research agenda for AI in the environmental sciences. *Risk Analysis* <https://doi.org/10.1111/risa.14245> (2023).
27. Cains, M. G. et al. Exploring NWS Forecasters’ Assessment of AI Guidance Trustworthiness. *Weather and Forecasting*. [In press]
28. Wirz, C. D. et al. NWS forecaster perceptions of new AI guidance for coastal fog prediction. In *Presented at the Joint Session with the 23rd Conference on Artificial Intelligence for Environmental Science and 19th Symposium on Societal Applications: Policy, Research and Practice at the American Meteorological Society 2024 Annual Meeting* (AMS, 2024).

29. Schmidt, A. H. et al. A new model for training graduate students to conduct interdisciplinary, interorganizational, and international research. *BioScience* **62**, 296–304 (2012).
30. Neal, J. W., Posner, S. & Brutzman, B. Understanding brokers, intermediaries, and boundary spanners: a multi-sectoral review of strategies, skills, and outcomes. *Evid. Policy* **19**, 95–115 (2023).
31. Council, N. R. et al. Recommendations for improving risk communication. In *Improving Risk Communication* (National Academies Press (USA), 1989).
32. Burke, A., Snook, N., Gagne, D., McCorkle, S. & McGovern, A. Calibration of machine learning-based probabilistic hail predictions for operational forecasting. *Weather Forecast* **35**, 149–168 (2020).
33. Sobash, R. A. et al. Diagnosing storm mode with deep learning in convection-allowing models. *Mon. Weather Rev.* **151**, 2009–2027 (2023).
34. Wirz, C. D. et al. (re)conceptualizing trustworthy AI: a foundation for change. *Submitted to Artificial Intelligence* (under review).
35. Dunbar, K. How scientists think: on-line creativity and conceptual change in science. In *Creative Thought: An Investigation of Conceptual Structures and Processes*. 461–493 (American Psychological Association, 1997).
36. Dunbar, K. Concept discovery in a scientific domain. *Cognitive sci.* **17**, 397–434 (1993).
37. Kamangir, H. et al. Fognet: A multiscale 3d cnn with double-branch dense block and attention mechanism for fog prediction. *Mach. Learn. Appl.* **5**, 100038 (2021).
38. Kamangir, H., Krell, E., Collins, W., King, S. & Tissot, P. Importance of 3d convolution and physics-based feature grouping in atmospheric predictions. *Environ. Model. Software* **154**, 105424 (2022).
39. Wirz, C. et al. Increasing the reproducibility, replicability, and evaluability of supervised ai/ml in the earth systems science by leveraging social science methods. *Earth Space Sci. (Submitted to Special Issue on Advances in Machine Learning for Earth Science: Observation, Modeling, and Applications)* (under review).
40. Fisher, A., Rudin, C. & Dominici, F. All Models are Wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *arXiv* <https://doi.org/10.48550/arXiv.1801.01489> (2018).
41. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
42. Kashinath, K. et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philos. Trans. A Math. Phys. Eng. Sci.* **379**, 20200093 (2021).
43. Holbach, H. M. et al. Recent advancements in aircraft and in situ observations of tropical cyclones. *Trop. Cyclone Res. Rev.* **12**, 81–99 (2023).
44. Ricciardulli, L. et al. Remote sensing and analysis of tropical cyclones: current and emerging satellite sensors. *Trop. Cyclone Res. Rev.* **12**, 267–293 (2023).
45. Harnos, D. S. & Nesbitt, S. W. Convective structure in rapidly intensifying tropical cyclones as depicted by passive microwave measurements. *Geophys. Res. Lett.* <https://doi.org/10.1029/2011GL047010> (2011).
46. Haynes, K., Slocum, C., Knaff, J., Musgrave, K. & Ebert-Uphoff, I. Aiding tropical cyclone forecasting by simulating 89-ghz imagery from operational geostationary satellites. In *AMS Conference on Hurricanes and Tropical Meteorology* (American Meteorological Society, 2022).
47. Kirchhoff, C. J., Carmen Lemos, M. & Dessai, S. Actionable knowledge for environmental decision making: broadening the usability of climate science. *Annu. Rev. Environ. Resour.* **38**, 393–414 (2013).
48. Rittel, H. W. & Webber, M. M. Dilemmas in a general theory of planning. *Policy sci.* **4**, 155–169 (1973).
49. McGovern, A. & Allen, J. Training the next generation of physical data scientists. *EOS* **102**, 1–9 (2021).
50. Morgan, M. G. *Interdisciplinary Research on Climate and Energy Decision Making: 30 Years of Research on Global Change* 1st ed, Vol. 336 (Routledge, 2023).
51. Bauer, P. et al. Deep learning and a changing economy in weather and climate prediction. *Nat. Rev. Earth Environ.* **4**, 507–509 (2023).

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. ICER-2019758. The authors thank Melissa Wilson Reyes for her comments on the paper.

## Author contributions

A.M., J.D., and A.B. shared the writing of the main manuscript and prepared Fig. 1. K.D.M. prepared Fig. 2. All authors reviewed the manuscript.

## Competing interests

The authors declare the following competing interests: Amy McGovern worked at Google Research as a visiting faculty member while this paper was written but none of her work at Google is related to this paper nor was any pay from Google used for this paper. The authors declare no other competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Amy McGovern.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024