

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



AI2ES Risk Communication research team



Julie Demuth



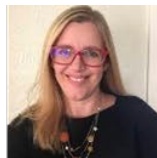
Ann Bostrom



Chris Wirz



Mariana Cains



Andrea Schumacher



Deiana Madlambayan
(graduated)



Jacob Radford



Susan Campbell



Erin Smith

Julie - training in atmospheric science and risk communication, focusing on hazardous weather experiences, risk perceptions, risk communication, and responses, among forecasters and publics.

Ann - training in risk and decision analysis and behaviors, research focus on mental models of hazardous processes and on risk perception and communication, e.g.,

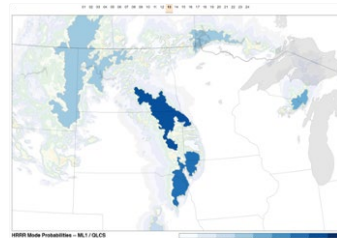
- How do people understand climate change?
- How do—and can—different approaches to communicating about climate change and risks of climate change affect and inform decisions about climate change?
- What drives trust in and trustworthiness judgments of AI/ML for weather and climate?

This material is based upon work supported by
the National Science Foundation
under Grant No. ICER (now RISE)-2019758

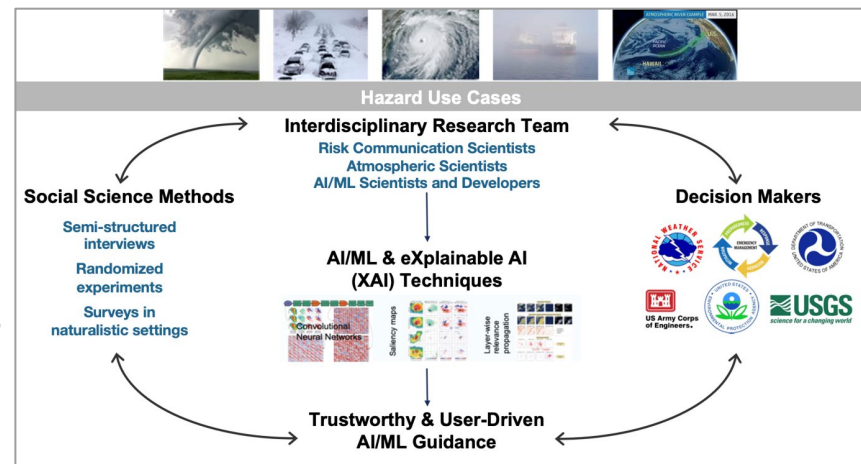


AI outputs as risk communication

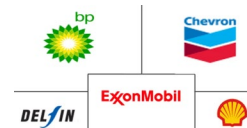
- **AI/ML model guidance is a form of risk information** for expert, professional users* who can use it to manage risks from weather, climate, and coastal hazards (whether they'll occur, when, how severe, etc.)
- **Trust and trustworthiness (which are social constructs) are of particular, focal interest** for developing and refining AI/ML model guidance
- We draw on theoretical and empirical social science literature and apply social science methods to investigate fundamental RC research goals



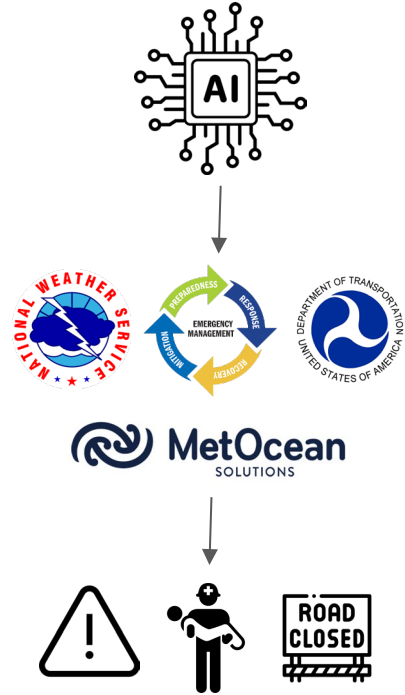
Credit: Ryan Sobash, David John Gagne, and others



* e.g., weather forecasters, transportation officials, emergency managers, oil and gas companies



AI trustworthiness perceptions of professional decision-makers

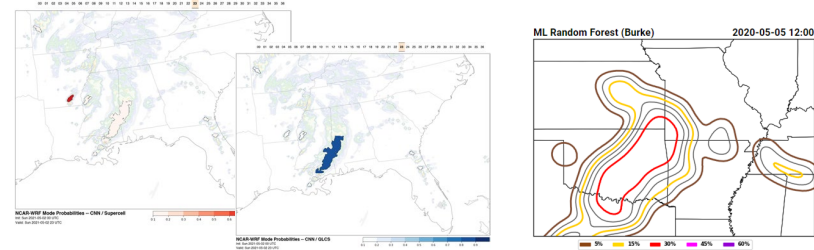


AI predictions function
as a type of risk
information

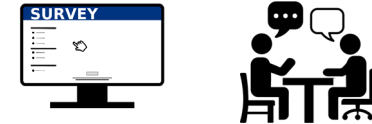
Used by professional
decision-makers

To assess risks of
weather hazards, make
critical job-specific
decisions

Used 2 new prototype AI models:
CNN probabilities of storm mode &
RF probabilities of severe hail

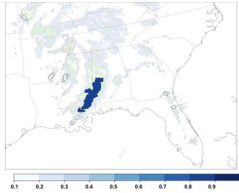


Conducted surveys & interviews with
forecasters

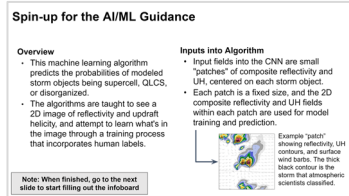


RQ: How do different attributes – e.g., AI technique used, the training of the AI model, the AI model input variables, AI model performance – influence forecasters' perceptions of model trustworthiness?

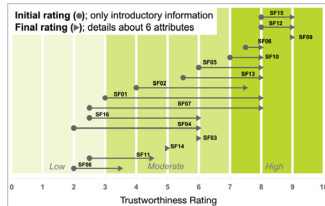
Findings at three “scales”



@ prototype “scale”: Developers hand-labeling inputs to develop predictions of storm mode increased forecasters’ trustworthiness if developers had relevant domain expertise. Thus, the **resource-intensive task of human hand-labeling may be important for some purposes.**



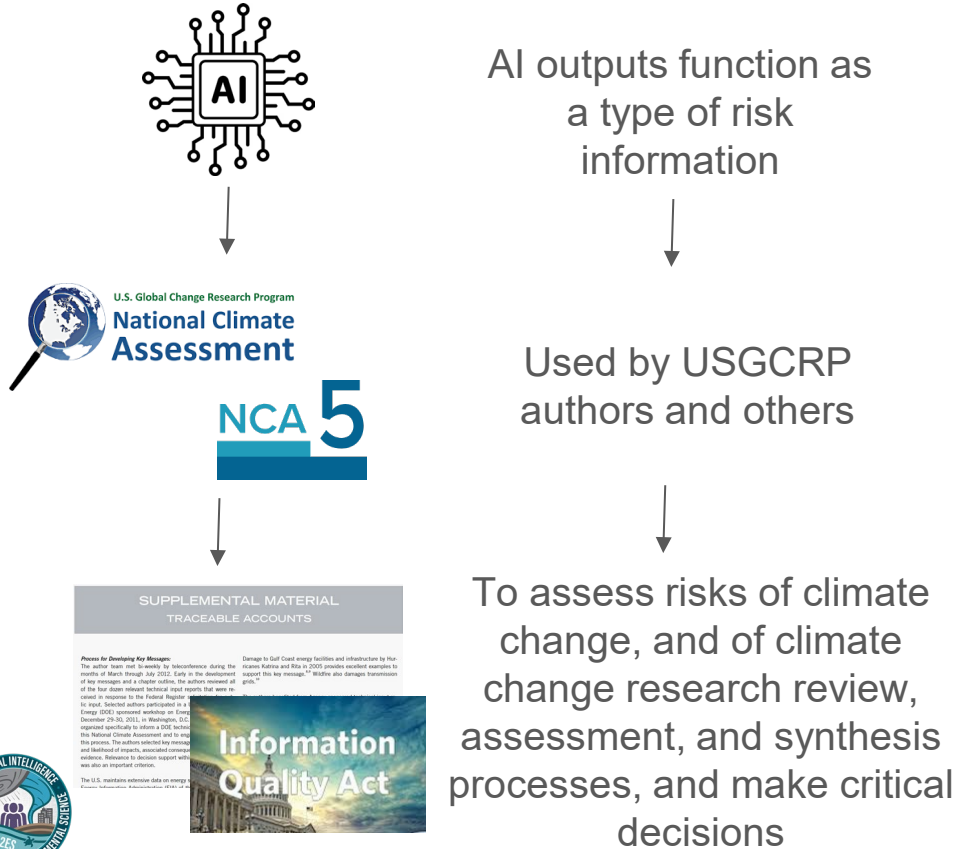
Across prototypes and attributes: forecasters’ trustworthiness = f (information about the AI model technique especially **input variables**, information about the **model performance especially failure modes**, **being able to interact** with the AI model output)



Overall: Forecasters’ trust in new AI guidance is a progressive process, not instantaneous and not maximized at outset



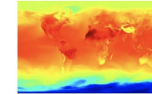
AI trustworthiness perceptions of USGCRP decision-makers



Existing and emerging prototype AI models

IBM announces new open-source AI model for weather and climate applications

Environment 2024 September 2024



SEMANTIC SCHOLAR



Claude

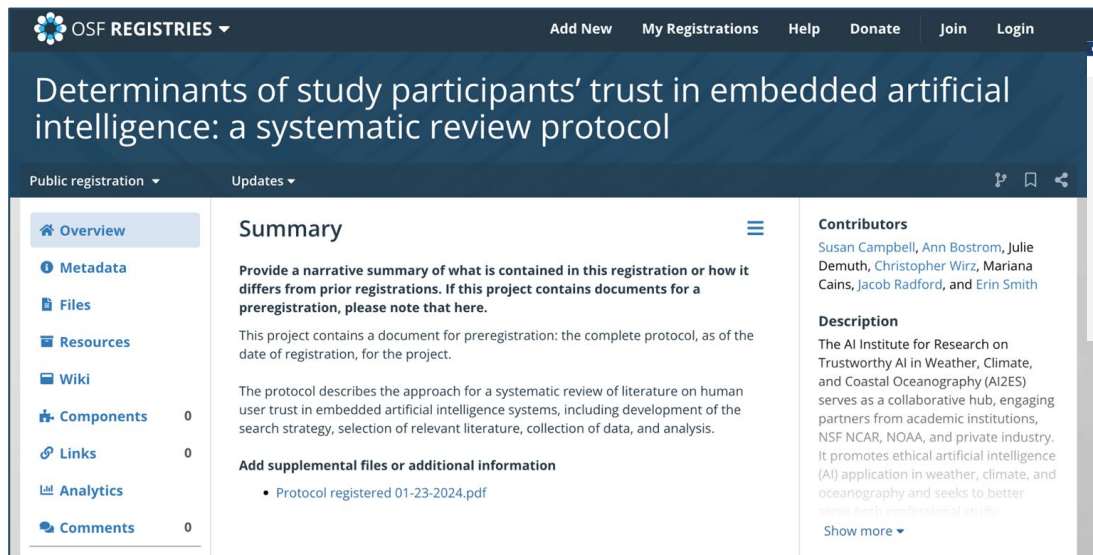
BY ANTHROPIC

Collaborate with US AI Safety Institute and others to conduct research with USGCRP authors



RQ: How do different attributes – e.g., AI technique used, the training of the AI model, the AI model input variables, AI model performance – influence reviewers' and writers' perceptions of model trustworthiness?

Risk communication (RC): New approaches to advancing research on trust in AI



The image shows the OSF Registries page for a pre-registered systematic review protocol. The title is "Determinants of study participants' trust in embedded artificial intelligence: a systematic review protocol". The page includes a sidebar with navigation links: Overview, Metadata, Files, Resources, Wiki, Components (0), Links (0), Analytics, and Comments (0). The main content area is divided into three sections: Summary, Contributors, and Description. The Summary section provides a narrative summary of the project, stating that it contains a document for preregistration of the complete protocol. The Contributors section lists Susan Campbell, Ann Bostrom, Julie Demuth, Christopher Wirz, Mariana Cains, Jacob Radford, and Erin Smith. The Description section states that the project is a collaborative hub for research on trust in AI, serving as a hub for academic institutions, NSF NCAR, NOAA, and private industry.

OSF REGISTRIES

Add New My Registrations Help Donate Join Login

Determinants of study participants' trust in embedded artificial intelligence: a systematic review protocol

Public registration Updates

- Overview
- Metadata
- Files
- Resources
- Wiki
- Components 0
- Links 0
- Analytics
- Comments 0

Summary

Provide a narrative summary of what is contained in this registration or how it differs from prior registrations. If this project contains documents for a preregistration, please note that here.

This project contains a document for preregistration: the complete protocol, as of the date of registration, for the project.

The protocol describes the approach for a systematic review of literature on human user trust in embedded artificial intelligence systems, including development of the search strategy, selection of relevant literature, collection of data, and analysis.

Add supplemental files or additional information

- Protocol registered 01-23-2024.pdf

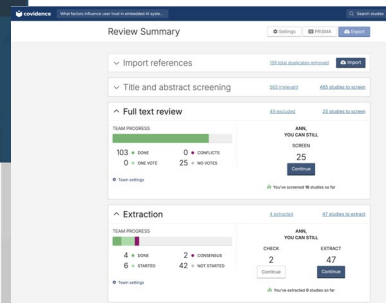
Contributors

Susan Campbell, Ann Bostrom, Julie Demuth, Christopher Wirz, Mariana Cains, Jacob Radford, and Erin Smith

Description

The AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES) serves as a collaborative hub, engaging partners from academic institutions, NSF NCAR, NOAA, and private industry. It promotes ethical artificial intelligence (AI) application in weather, climate, and oceanography and seeks to better serve both professional study

Show more



The image shows the Covidence Review Summary interface. It displays a progress bar for the review process, with sections for Import references, Title and abstract screening, Full text review, and Extraction. The Full text review section shows a progress bar and a table of results. The Extraction section shows a progress bar and a table of results.

Covidence

Review Summary

Import references

Title and abstract screening

Full text review

Extraction

In progress:

- Systematic review of research on trust in embedded AI developed and pre-registered on OSF (2024): Susan Campbell, Ann Bostrom, Julie Demuth, Christopher Wirz, Mariana Cains, Jacob Radford, and Erin Smith, *Determinants of study participants' trust in embedded artificial intelligence: a systematic review protocol*, osf.io/6mwgz
- Working in Covidence, we have completed all title and abstract screening of papers identified through Web of Science (2010 through 2023), and are in the process of full text extraction and critical appraisal of the 62 relevant papers.



Emerging findings, questions for USGCRP

- **Trustworthiness stems from intersecting factors**, including: users' decision-making needs and contexts; data quality and representativeness; model development processes, techniques, and specifics; model availability, interpretability, explainability, and integration into users' workflows; perceptions of the model developers' expertise; and **model skill** (i.e., performance) across hazards and geography.
- **Trust is inherently emotional and subjective; this complicates efforts to “calibrate” trustworthiness.**
- **To develop trustworthy AI/ML:**
 - Improve measurement of trust in AI as a dynamic, contingent process
 - Learn which contingencies and contextual factors matter – through co-design/co-production and engagement across the entire AI lifecycle
 - Develop and test strategies for communicating the uncertainties of AI/ML model outputs
- **To consider:**
 - Assessing how AI is used in USGCRP work as well as its inputs.
 - Rapid evolution of AI - the need for agile approaches to assessment.

