Successes, Challenges, and Collaboration Areas for AI in ESP at NSF

Eric DeWeaver
Program Director
Climate and Large-scale Dynamics
Atmosphere Section
Atmospheric and Geospace Sciences Division, GEO

4 November 2024

AI and ESP at NSF

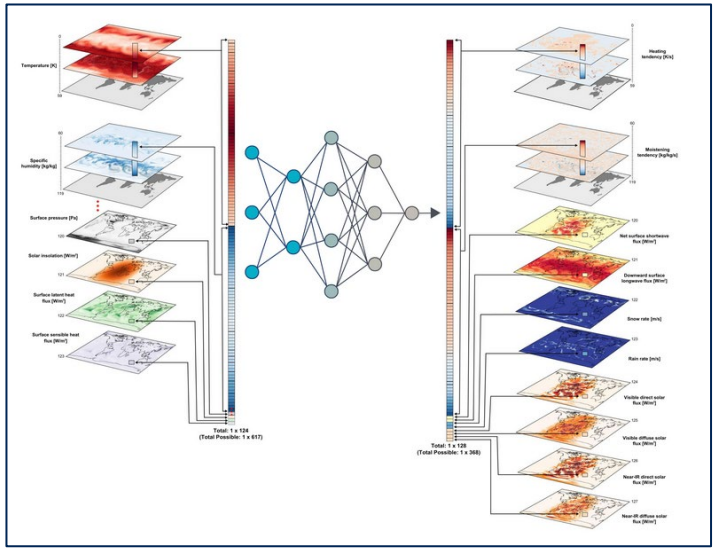
- NSF supports basic science research to develop better understanding of the earth system, including its predictability.
- We value earth system models, in particular CESM, primarily as tools for basic science research rather than as models for making predictions.
- We seek to derive earth system understanding from AI, thus we're interested in what ML can teach us about the earth system, including its predictability.
- We have a strong interest in climate change, which has different priorities from weather prediction: probabilistic/risk framework, enforcement of conservation laws, generalization beyond the distribution, attribution to forcing vs internal variability, etc.
- NSF supports basic research on AI/ML methods through its math (MPS) and computer science (CISE) programs.
- Workforce development is an important part of our mission.

AI “successes” at NSF

LEAP: The Center for Learning the Earth with Artificial Intelligence and Physics

- Science and Technology Center at Columbia University
- Seeks to integrate ML into the Community Earth System Model (CESM) to improve climate projections.
- Also developing the discipline of Climate Data Science, building on Pangeo interface.
- Partners include NSF NCAR (CESM developers), NASA/GISS, and NVIDIA.

ClimSim

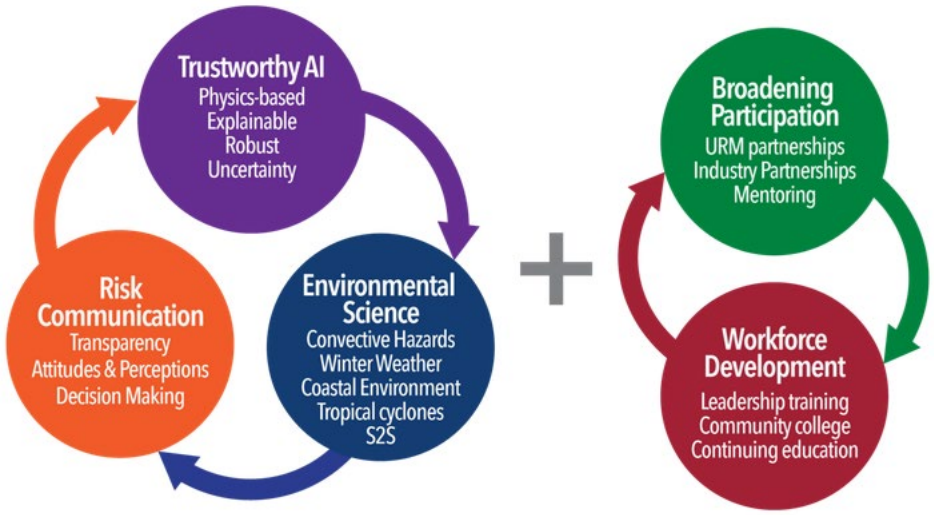


A Kaggle competition to create column physics emulators (LEAP + many partners)

AI “successes” at NSF

AI2ES: AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography

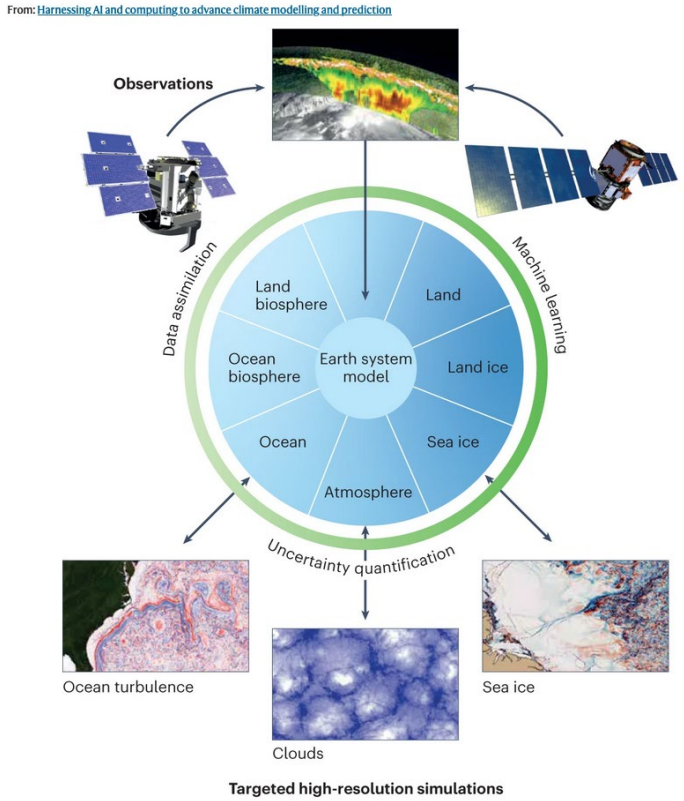
- AI Institute based at the University of Oklahoma
- Develops and tests trustworthy AI methods for environmental prediction
- Partners include NOAA, NSF NCAR, NVIDIA, Google, and The Weather Company
- Emphasis on workforce development



AI “successes” at NSF

CLiMA: The Climate Modeling Alliance

- Project to build a hybrid Physics/ML ESM
- Based at CalTech, sponsors include Schmidt, CISCO, DoD, NSF
- Development of emulators for automated parameter tuning
- Use of nondimensional groups to create ML schemes robust to climate change
- Software innovations for incorporating ML into ESMs and scaling on GPUs



AI “successes” at NSF

CAIG: Collaborations in AI and Geoscience

- New joint solicitation (NSF GEO, CISE, and MPS Directorates) focused on establishing meaningful collaborations between geoscientists and AI experts
- Goals: Advance innovative use and application of AI in geoscience; Expand access to data and computation for AI-driven geosciences; Build workforce capacity through educational and broadening participation activities
- Currently supporting multiple projects doing AI for ESP
- Democratizes AI research by supporting academic researchers who are not part of a big center

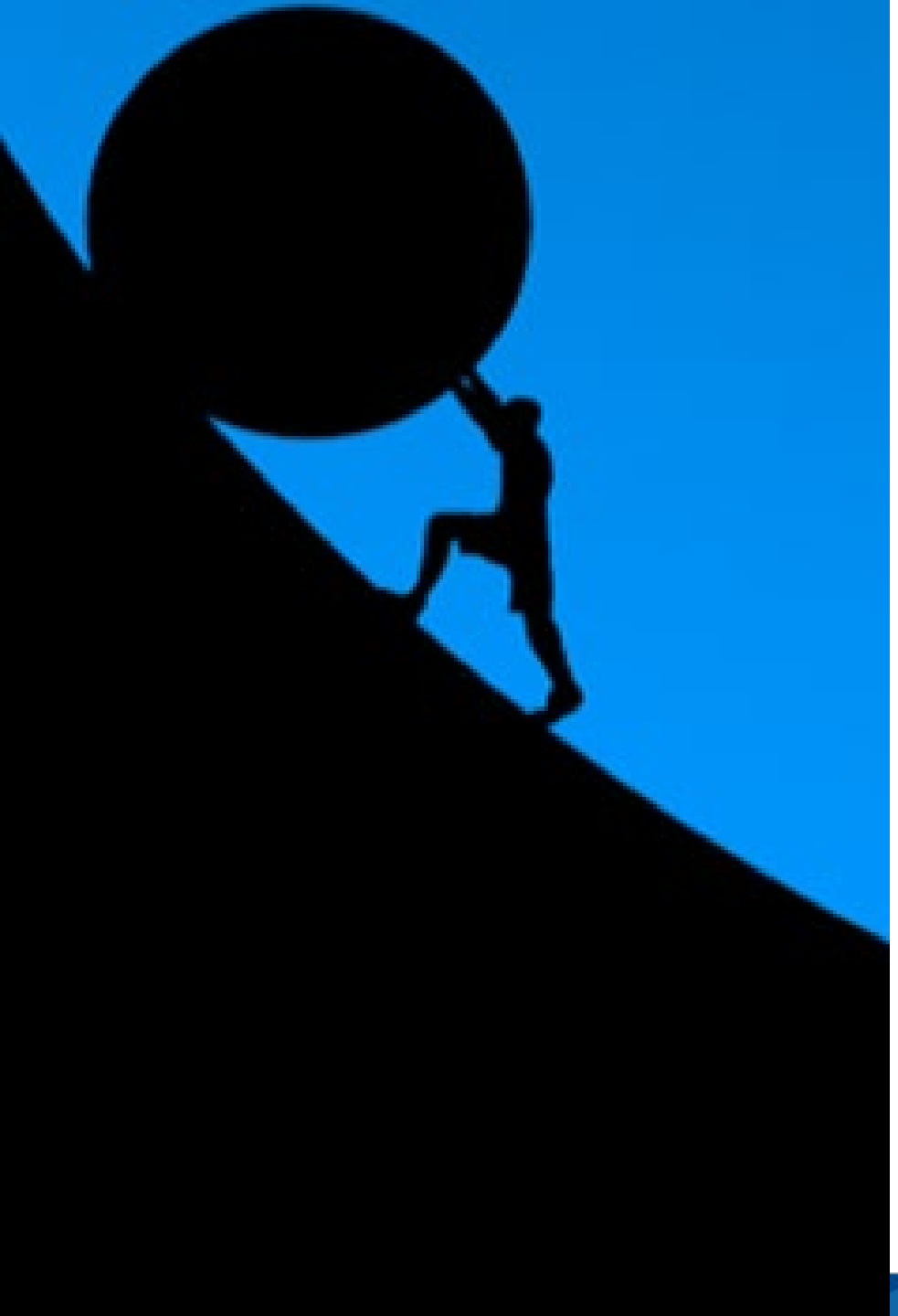
National Discovery Cloud for Climate (NDC-C)

NSF initiative for a national-scale cyberinfrastructure to support end-to-end climate research and education

Pilot funding provided to **31 projects** including support data-intensive workflows for AI and climate, e.g.,

- Integrating NSF NCAR data infrastructure with the Open Science Data Federation
- The National Data Platform Pilot (UCSD) is investigating data discovery, federation and usability for AI and climate
- Allocations and training for climate-related AI/ML computing via DeltaAI (NCSA)

Projects are exploring mechanisms for achieving federated, collective impact



Challenges for AI and ESP

- Access to training datasets
- Access to cloud and HPC resources
- Porting/software for GPUs
- APIs to connect Fortran to ML libraries
- FAIR data principles, incl. data licensing issues
- Workforce training
- Unfamiliar communities and cultures
- Lack of peer review (e.g. ArXiv, conf. proc.)

Potential Areas for Collaboration

- Data access – including reanalysis products, satellites data, PCMDI/ESG, LES/GCRM simulation libraries
- Benchmark datasets
- Foundation models
- Software for porting codes to GPUs
- APIs for connecting legacy codes to ML libraries
- Computing resources, e.g. cloud computing
- Activities through the NSF NAIRR Pilot, mandated by EO 14110

Satellite data sharing for AI/ML systems

Data-driven models are data hungry: they work when all data from all sources can be freely accessed and utilized.

License restrictions are a killer for this approach.

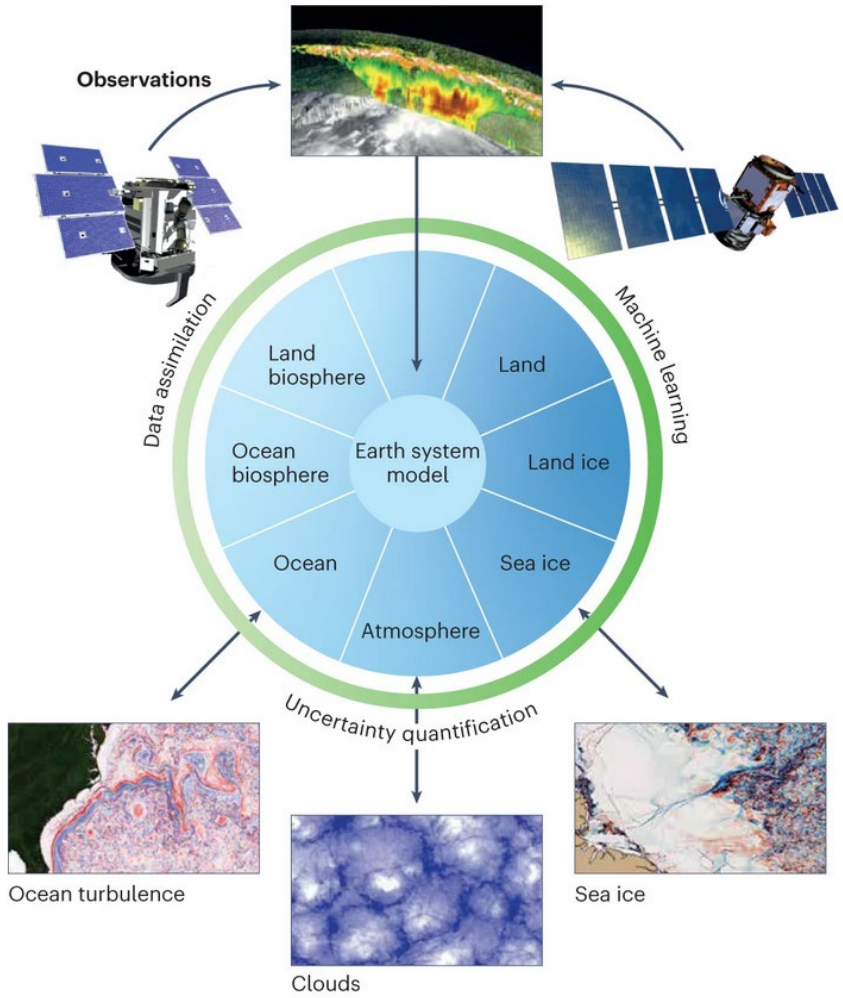
This is an issue for satellite data as Congress mandates commercial data buys:

Congress “encourages NOAA to … refrain from building systems that compete directly with industry”.

As purchased data replaces government missions license issues will become more prominent.

ICAMS should address this issue.

From: Harnessing AI and computing to advance climate modelling and prediction



Final Thoughts

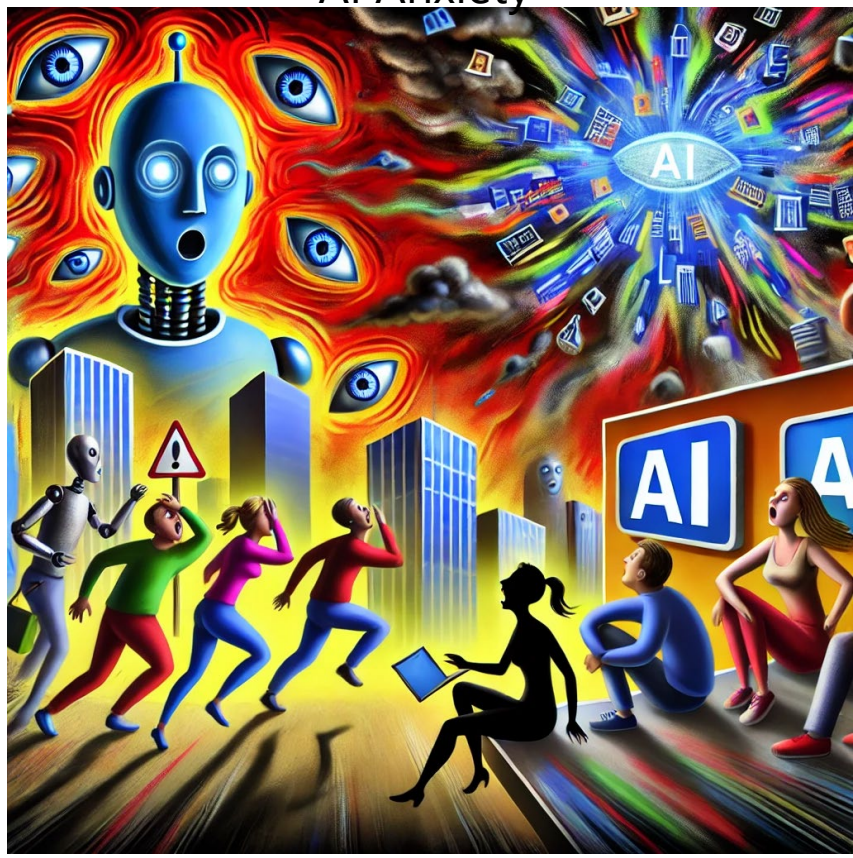
There is great excitement about the promise of AI, in ESP and everything else.

But: there's also great anxiety about AI:

- We don't know how it works, what it can do, what to expect.
- We afraid of falling behind.
- The Government wants AI research NOW!

Any help you can offer will be much appreciated.

AI Anxiety



ChatGPT (pers comm, 2024)