Computational Geosciences (2024) 28:1305–1329
https://doi.org/10.1007/s10596-024-10317-7

ORIGINAL PAPER

Towards practical artificial intelligence in Earth sciences

Ziheng Sun[1] · Talya ten Brink[2] · Wendy Carande[3] · Gerbrand Koren[4] · Nicoleta Cristea[12] · Corin Jorgenson[14] · Bhargavi Janga[1] · Gokul Prathin Asamani[1] · Sanjana Achan[1] · Mike Mahoney[5] · Qian Huang[6] · Armin Mehrabian[7] · Thilanka Munasinghe[8] · Zhong Liu[1,7] · Aaron Margolis[10] · Peter Webley[13] · Bing Gong[11] · Yuhan Rao[9] · Annie Burgess[16] · Andrew Huang[17] · Laura Sandoval[3] · Brianna R. Pagán[7] · Sebnem Duzgun[15]

Received: 20 November 2023 / Accepted: 8 August 2024 / Published online: 2 September 2024

Abstract
Although Artificial Intelligence (AI) projects are common and desired by many institutions and research teams, there are still relatively few success stories of AI in practical use for the Earth science community. Many AI practitioners in Earth science are trapped in the prototyping stage and their results have not yet been adopted by users. Many scientists are still hesitating to use AI in their research routine. This paper aims to capture the landscape of AI-powered geospatial data sciences by discussing the current and upcoming needs of the Earth and environmental community, such as what practical AI should look like, how to realize practical AI based on the current technical and data restrictions, and the expected outcome of AI projects and their long-term benefits and problems. This paper also discusses unavoidable changes in the near future concerning AI, such as the fast evolution of AI foundation models and AI laws, and how the Earth and environmental community should adapt to these changes. This paper provides an important reference to the geospatial data science community to adjust their research road maps, find best practices, boost the FAIRness (Findable, Accessible, Interoperable, and Reusable) aspects of AI research, and reasonably allocate human and computational resources to increase the practicality and efficiency of Earth AI research.

Keywords Artificial intelligence (AI) · Machine learning (ML) · Earth sciences · FAIR · MLOps · Practical AI

✉ Ziheng Sun
zsun@gmu.edu

✉ Talya ten Brink
talya.tenbrink@noaa.gov

1 Department of Geography and Geoinformation Science, Center for Spatial Information Science and Systems, George Mason University, Fairfax, USA

2 Greater Atlantic Regional Fisheries Office, NOAA Fisheries, Gloucester, USA

3 Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, USA

4 Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands

5 SUNY-ESF, Syracuse, USA

6 Center for Rural Health and Research, East Tennessee State University, Johnson City, TN, USA

7 NASA Goddard Earth Sciences Data and Information Services Center, Greenbelt, MD, USA

8 Rensselaer Polytechnic Institute (RPI), Troy, USA

9 North Carolina Institute for Climate Studies, North Carolina State University, Raleigh, USA

10 Environmental Protection Agency, Washington D.C., USA

11 Jülich Supercomputing Centre, Jülich, Germany

12 eScience Institute, University of Washington, Seattle, USA

13 University of Alaska Fairbanks, Fairbanks, USA

14 Faculty of Science, University of Geneva, Geneva, Switzerland

15 Colorado School of Mines, 1500 Illinois Street, Brown Hall 268, Golden CO, USA

16 Montana State University, Bozeman, MT, USA

17 Anaconda Inc, Austin, USA

1 Introduction

Recently, the rapid growth of artificial intelligence (AI) has ushered in transformative possibilities across various domains, including Earth sciences. While AI holds immense potential to revolutionize how scientists collect, analyze, and interpret data, there exists a significant hurdle to its practical application. The current landscape of AI in Earth sciences presents formidable challenges, requiring substantial efforts that may hinder widespread engagement in AI-driven research. Despite exciting lab results in mineral exploration, seismic analysis, and climate modeling, the practical implementation of AI in these fields demands meticulous attention [1–3]. We define practical AI as AI that is used in real-world applications or has strong potential for real-world use. This includes AI technologies that have been tested and validated in relevant environments and are intended for practical deployment. For more context, practical AI corresponds to Level 3 or higher in the NASA Technology Readiness Scale [4], which indicates that the technology has been validated in relevant environments and shows promise for practical implementation. One of the most compelling facets of AI in Earth sciences is its capacity to handle vast datasets. For instance, satellite imagery can detect changes in land use, monitor environmental degradation, and study natural disasters. In mineral exploration, AI algorithms can identify mineral deposits based on geophysical data patterns, minimizing the need for extensive field surveys. Additionally, AI enhances seismic data analysis, aiding geologists in understanding Earth's interior structure and predicting earthquakes [5, 6]. Climate modeling, a complex and time-consuming process, benefits from AI techniques, such as machine learning and deep learning, improving accuracy and efficiency [7, 8].

Amidst these advancements, popular general AI applications like ChatGPT for writing and Midjourney for image creation represent noteworthy progresses that science community can learn from. These applications showcase AI's potential to significantly impact geoscientific practices by uncovering relationships within diverse datasets and unveiling patterns that were previously hidden. The dynamic application of AI in Earth sciences signifies a rapidly evolving and exciting field. Its ability to analyze extensive datasets, streamline tasks, and reveal concealed relationships positions AI to revolutionize geoscientific methodologies, facilitating once-unattainable discoveries. This paper endeavors to explore best practices and future directions, providing insights to enhance the practicality and usability of AI for Earth scientists. Collaboration across disciplines, including computer science, mathematics, and geology, is imperative to develop AI tools tailored for Earth sciences. Educating and training geoscientists in AI tool usage fosters a culture of innovation and collaboration. Addressing these factors is crucial to making AI more practical and usable, potentially leading to groundbreaking discoveries. The included roadmap for AI (as shown in Fig. 1) elucidates its evolution from specialized training to integration into diverse datasets, reinforcing the importance of ongoing learning, monitoring AI behavior, and navigating ethical considerations. This paper delves into the existing challenges, best practices, and future

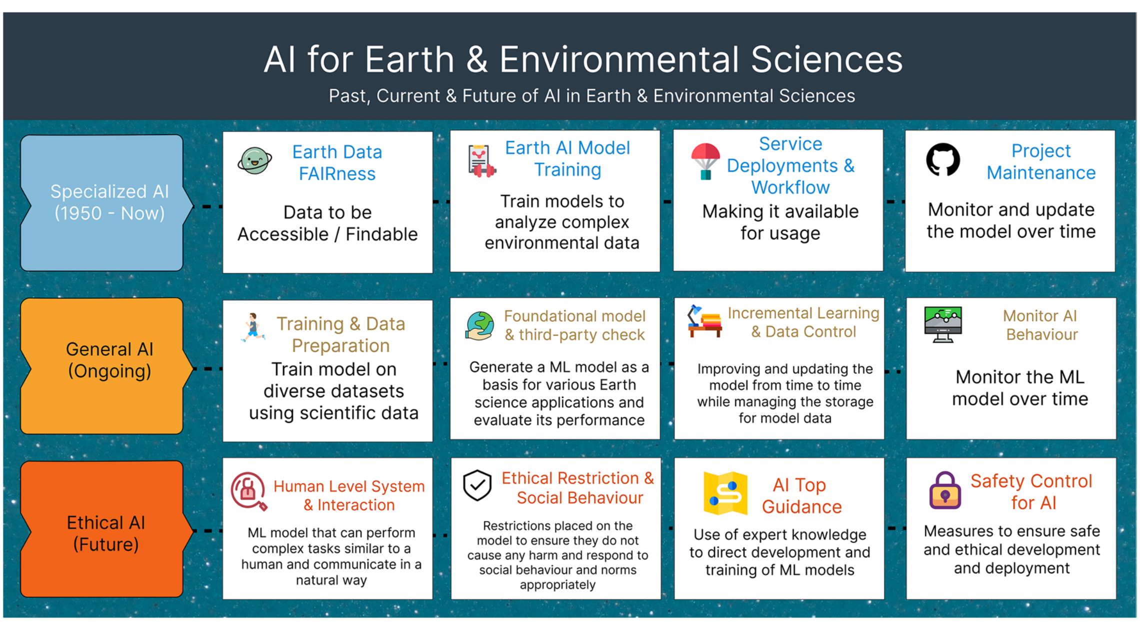


Fig. 1 Road map for AI applications in Earth sciences

aspirations of AI in Earth sciences, recognizing the necessity of a collective effort to unlock its full potential.

2 The importance of AI in the Earth sciences

Before putting AI into practice, we need to understand what AI can do for Earth sciences. In other words, what will the future Earth scientific research look like? This section introduces research directions and attempts to picture what practical AI products or services will look like in the future.

2.1 Data collection and processing

Much of the work for data collection and processing has been automated (Fig. 2). Coordinated data collection, standardization, and open data sharing can facilitate scientific research on large scientific problems, for example, global environmental change [9], which can be further accelerated by AI-approaches. Our future society will continue to rely significantly on the current or under-development data infrastructure, like satellites, drones, stations, in-situ sensors, mobile devices, etc. AI is expected to help augment the capture and processing of daily or on-premise datasets. For example, due to the interruption by uncontrollable variables like sun magnetic storms, sensor accidental malfunctions, clouds, extreme weather, dead batteries, etc., there is always missing data and bad quality data. AI has been one of the promising solutions to deliver seamless time series by automated gap-filling. One typical machine learning application is fixing the Landsat 7 imagery with stripes because of the failure of Scan Line Corrector since 2003 (reference). In the future, we can expect AI services that can intelligently fill and adjust the originally collected raw data to create more complete and continuous observations, which is always ideal. AI-enhanced data enrichment can increase meaningful and actionable information for scientists from abundant data and provide a firmer bond between science and society.

Ideally, techniques like Diffusion Models [10] and GAN (Generative Adversarial Networks) [11] could create reliable data based on other variables' data series, even if there is no device actually observing that variable. This action will save a huge number of resources and avoid deploying function-overlapping physical sensors. For example, the future Earth Science Community could deploy a single stationary network to collect all the fundamental datasets, and scientists across the spectrum of Earth Science domains can derive their domain-specific datasets from using AI. We can reuse the existing satellites or launch a new series of satellites to form a constellation covering the globe with a short revisit time and rich radio spectrum. Then produce all the datasets from the raw satellite observations using AI services in an automated manner. Even if the original constellation did not meet the coverage or frequency requirements by the domain, or even did not cover the domain in the original proposals, AI can model the relationships and transform the dataset to new datasets that are directly needed by scientists from the new domain.

2.2 Anomaly detection

Anomalies refer to those events that do not fit into the expected pattern or the known physics of a model and is one of the most important pieces of information for scientists and stakeholders. Detecting anomalies in the sea of big data is a key task for AI/ML in the era of big data. Future Earth science communities will begin to understand the interconnected or teleconnected processes in Earth systems to get a full picture of the underlying mechanisms. Right now, scientists are challenged by the need to single out anomalies. It has been a very challenging task to distinguish useful anomalies from data noises or non-meaningful events [12, 13].

An example of how anomalies can be detected using ML is by using a hybrid architecture combining deep belief networks (DBN) and one-class support vector machines (OCSVM) [14]. The DBN model is used to extract abstract features, which are then fed into the OCSVM for anomaly detection. The DBN model's training happens layer-by-layer, which enables the extraction of relevant features from the input data. When dealing with an input vector vi, the activation of the hidden units h_i is determined as follows:

$$h_i = f\left(b_i + \sum_j w_{ij} v_i\right) \tag{1}$$

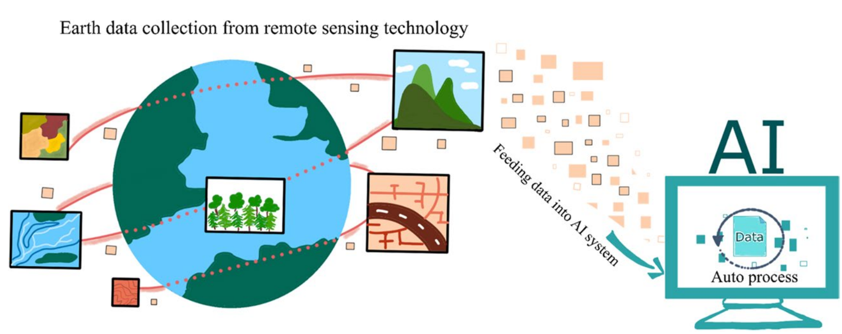


Fig. 2 Data collection and processing illustration

where f is the sigmoid function, b_i biased values, and w_{ij} the weight matrix. The extracted features are thus provided as inputs to OCSVM, which finds a hyper-plane boundary to effectively separate the multivariate anomaly from the background. The decision function f(x), used for determining hyper-plane is expressed as:

$$f(x) = sin(w * \phi(x) - \rho) \tag{2}$$

where w is the weighted vector expressed as,

$$w = \sum_{i=1}^{m} a_i \phi(x_i) \tag{3}$$

where $\phi(x)$ is the map function and ρ is the offset.

In the context of anomaly detection, the hyper-plane determined using the above decision function can be considered an indicator to assess whether a sample, denoted as x, is classified as an anomaly or not based on the following set of rules:

$\mathrm{f(x)} > 0$, if x ϵ background

$\mathrm{f(x)} < 0$, if x ϵ *anomaly*

The hyperplanes created through the combination of deep belief networks (DBN) and one-class support vector machines, both of which are ML methods, can be used for anomaly detection in geoscience. However, the field is changing fast [15] and a lot of new technologies are emerging quickly [16] since the ChatGPT was first released [17], e.g., large language models (transformer) [18, 19] and diffusion models [20]. Large language models have shown promise in anomaly detection tasks by leveraging their ability to understand and analyze textual data, enabling them to identify anomalies in natural language patterns or textual data streams. Similarly, diffusion models, known for their capability to generate high-quality images and understand complex data distributions, are now being explored for anomaly detection in image-based or video applications.

In the future, AI is expected to simplify the task by accurately and automatically detecting useful anomalies with prebuilt production-level AI systems. As anomaly detection is often directly connected to the alert systems (e.g., flooding alert, wind storm alert, etc.), the adoption of AI services may significantly reduce or eliminate alert spam or false alarms (Fig. 3). AI will relieve scientists from being overwhelmed by tedious data filtering tasks and focusing on finding solid evidence to answer core scientific problems. Another key place that AI could improve is the threshold settings. Right now, most threshold setting for anomalies is manually done by experts, which requires years of experience to find reasonable threshold values, which are generally static and may not be ideal in some time-sensitive cases (e.g., missing signals of early warning of landslides or wildfires) [21]. AI can dynamically adjust the thresholds based on sophisticated contexts and the knowledge AI models has learned from the accumulated decades of historical records, which will be more accurate and quicker than experts-adjusted threshold setting approaches. The expected results will allow earlier and more accurate alerts for all kinds of natural hazards and provide better opportunities for emergency response teams to act and contain the damages.

2.3 Monitoring and measurement

One of the main benefits brought by AI to Earth and environmental system monitoring is automation. Most teams want to involve AI to maximally reduce the level of time-consuming human intervention in their monitoring routines. Unmanned monitoring is more scalable to cover larger areas at a higher frequency (Fig. 4). With the capability of smoothly connecting workflow by direct data transformation and rule enforcement by AI, the latency between the observation time and the monitoring time by scientists is expected to be further condensed until it is close to real-time synchronization. Many scientists envision how AI could boost data quality in both temporal and spatial dimensions [22]. AI is intended to function as a guardian to block or limit poor quality data and only allow good quality data to initially reach the dashboard of scientists or decision makers.

The measurement strategy needs to be optimized for the targeted problems. For field measurements, it is troublesome for scientists to determine the best places to take measurements, how much battery the devices should be equipped with, and the interval for each observation period. AI techniques such as reinforcement learning can serve as

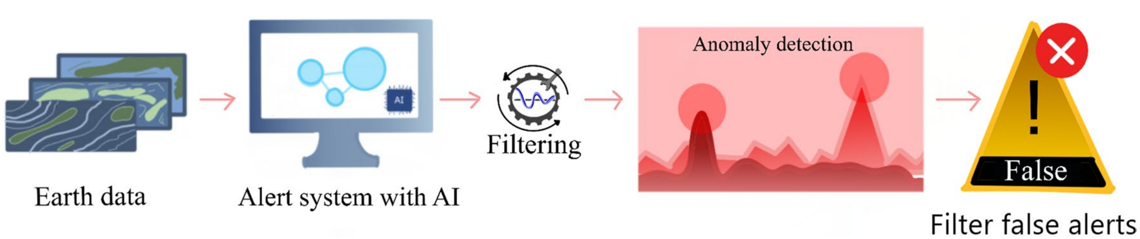


Fig. 3 Anomaly detection illustration

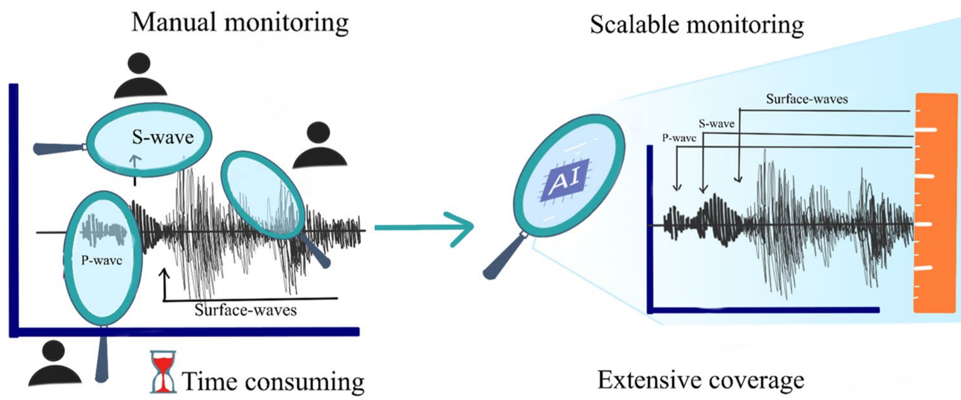


Fig. 4 Seismic wave monitoring and analysis with AI

optimization models to answer planning questions, such as the best route, or the best observations to achieve a research goal. The models can learn based on the rules and targets set by scientists, like the extent of valid places suitable for sensor planting, the target observation coverage, and the maximum number of available devices, etc. With algorithms like genetic algorithms, AI can turn the task into an optimization problem and find a reliable model to guide scientists to fulfill their measurement plan at low costs.

In addition, AI can create a pathway for blending those non-conventional monitoring approaches with the current standard monitoring strategy, such as crowdsourcing or citizen science projects. For many research teams, crowdsourcing is the most economical approach for monitoring and measurement collection. However, it is known that crowdsourced data quality is a big issue [23]. Although there are ways to improve data quality, such as assigning each crowdsourcer a "reliability score" and weighting the data from people with higher scores more heavily, it still poses a concern among data users regarding the overall quality. In the future, we expect that AI can be embedded into the collection devices by citizen scientists to guide them to take better quality observations. Meanwhile, AI services will be developed to boost the crowdsourced data quality and make them more usable and trustworthy by the science community.

2.4 Short-term prediction

Short-term prediction generally refers to the prediction made several hours or several days in advance, and is the most common prediction we receive on a daily basis and essential for social sectors like agriculture and aviation to function. Most weather services are short-term, including both hindcasting (within ~ 6 h) or forecasting (several days). Google-owned DeepMind already delivers improved short-term weather forecasting using AI models [24]. Many workshops have held to discuss how to use latest AI techniques in operational weather forecasting [25–27]. These workshops bring together experts from academia, industry, and government agencies to share insights, collaborate on research projects, and address challenges in applying AI to weather prediction. In the near future, we expect to see the adoption of AI becoming more common in production-level weather services. There are many AI companies and tech giants actively working toward that goal. For the general public and local communities, they will see more accurate and timely short-term weather forecasts because AI can save the computational-expensive calculation required by high-resolution predictions for short turnaround times (i.e. < 2 h). Generative models have been used to make nowcasting rainfall using the radar data hours earlier and the results are very promising compared to other existing models [24]. Because the short-term prediction cares more about the trends and there is limited time for the trends becoming unrecognizable, AI is considered to have a huge advantage to tackle the task, and provide probabilistic improved values with enhanced accuracy and time advance.

For the Earth science research communities, the involvement of AI is no doubt a huge transition from the traditional physics-informed numeric models to primarily data-driven AI models. Scientists will find the AI prediction less interpretable than numeric models as AI directly learned all the patterns from the data instead of pre-fixed physics equations. However, AI approaches can also strengthen traditional process-based models by effectively uncovering previously unknown relations between variables or processes as was e.g. shown in the field of ecology [28]. The reasoning within models and simulated processes will no longer be as transparent and adjustable as the traditional models. The primary focus of research will be slowly shifted from model parameter tuning to data engineering. However, in the future, AI may completely replace the existing numerical model-based prediction. Scientists will continue to work in a hybrid environment where numerical models and AI models coexist for a long time, and their relationship will be interdependent. An ideal but very possible situation may be that AI models will rely on numerical model results for training, while numerical models can use AI models to skip some computation-expensive steps.

It took several years before the scientific community picked up ML as a new approach and explored it in environmental applications. Instead of replacing the entire numerical weather prediction (NWP) models, the science community has explored how to use this new technology and smoothly evolution in each core of the current NWP workflow from data assimilation, forecasting to postprocessing. A review from [29] includes a thorough discussion on the possibility of replacing the core part of the NWP models and the opportunities and challenges of AI for weather and climate prediction. Recently, various types of ML and hybrid models applications are explored from replacing specific parameterizations in a model to speed up complex and time-consuming components [30] to uncertainty quantification [31] and improving models with post processing (e.g. downscaling; Price 2022). Table 1 listed the common research directions in this domain.

2.4.1 AI-enhanced physics model

AI-enhanced physics models, achieved through the creation of hybrid physics-AI models involve the incorporation of AI models into existing physics models (Fig. 5). AI models can replace one or more components of a physics-based model or predict intermediate quantities that are inadequately represented by physics alone. For instance, in a study by [32], recurrent neural network layers were integrated into the existing physics model to create a hybrid model that combines the strengths of both physics model and AI model. It constructs a deep learning architecture that incorporates these recurrent neural network layers into the existing physics model, allowing them to capture the temporal data of the physical system being modeled and map its temporal evolution to state variables. This enables the hybrid model to understand how the system changes over time, which is crucial for predicting complex physical phenomena. The AI components of the model are then utilized to extract the most significant features from the rich temporal data. The hybrid model is then trained using the extracted data. This resulting model excels at accurately predicting the complex behavior of physical phenomena [32]. The NN model and LSTM model rely solely on data-driven approaches, while the physical NN model is rooted in physics-based methods. In this comparative analysis of the four models, the hybrid model consistently outperforms the others, exhibiting superior accuracy.

2.4.2 Physics-informed AI model

Physics-Informed Neural Networks (PINNs) is a class of deep learning algorithms that integrate data with

Table 1 A comparison of the various ways of integrating AI and physics models

	Advantages	Disadvantages
Pure Physics Model	Predict long-term behaviors of damages (Disasters) and can be improved by comparing with observation using statistical methods like hypothesis testing and the Bayesian method [34]	Contain many assumptions and approximations and require model validation to ensure accuracy. The number of model parameters increases as model complexity increases, making it difficult to identify the model
Pure AI Model	Handle vast and intricate datasets and continuously enhances predictions by learning from experience and adapting to new data [35]	Could be largely influenced by unique types of noise and missing values in geoscience data and managing the spatial and temporal heterogeneity of data
Physics-informed AI	It can integrate the physical formulas directly into the network architecture; rather than solely depending on data, they can approximate and learn solutions to PDEs with high accuracy, even when the solution is discontinuous and noisy [36]	Training models can be time and resource intensive. The process to achieve a desirable level of accuracy can be computationally expensive and time-consuming
AI-enhanced Physics Model	Can improve predictions, enable the replacement of poorly modeled components of a physics-based model, capture dynamics that are otherwise absent from the model, and achieve better parameterization [32]	In the current literature, there exists a relatively limited number of AI-enhanced physics models under development, and not much existing work to learn from

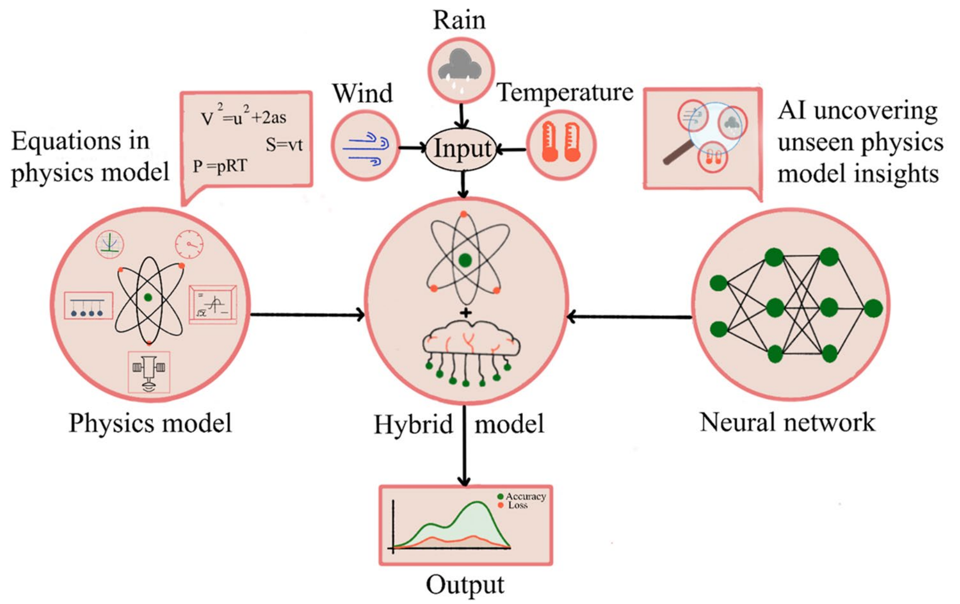


Fig. 5 Flowchart for representing the AI-enhanced physics model

mathematical operators and physics constants, including partial differential equations (PDEs) (Fig. 6). They are motivated by the need for machine learning methods that can handle imperfect data, such as missing or noisy values and outliers, and still provide accurate and physically consistent predictions.

One illustrative example as described in [33] on how PINNs integrate information from both measurements and partial differential equations (PDEs) by embedding the PDEs into the loss function of a neural network using automatic differentiation. The one-dimensional advection–diffusion equation, which encapsulates the behavior of scalar quantities such as temperature or moisture in the atmosphere, can be expressed as:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - v\frac{\partial^2 u}{\partial x^2} = 0 \tag{4}$$

where u is the representation of neural network, t is the time, and x represents the input variables. The loss function includes a supervised loss of data measurements of u from the initial and boundary conditions and an unsupervised loss of PDE described above are represented using the equation below:

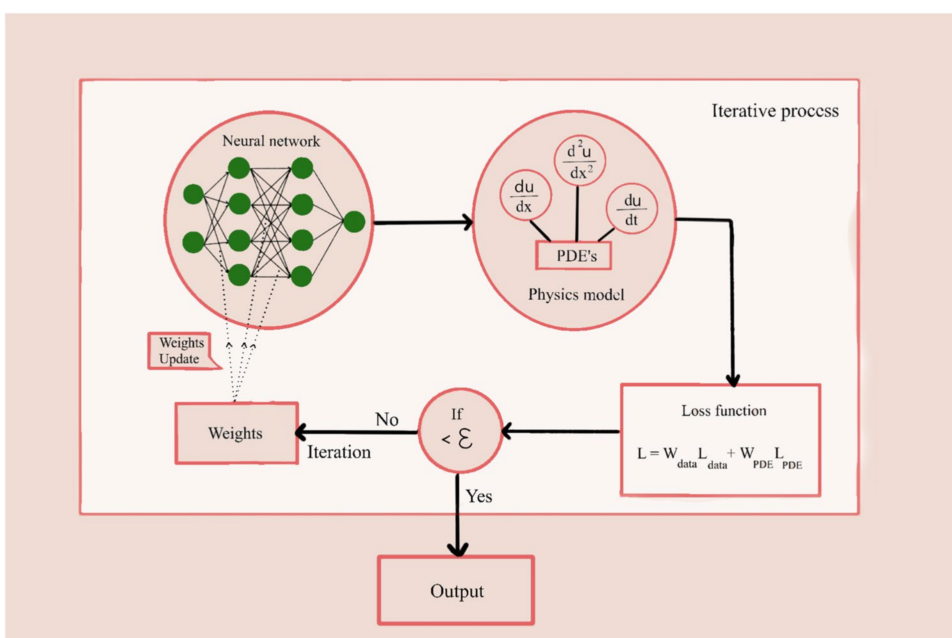


Fig. 6 Flowchart for representing physics informed AI model

$$L = w_{data}L_{data} + W_{PDE}L_{PDE} \tag{5}$$

where L is the loss function. The definition of L_{data} is:

$$L_{data} = \frac{1}{N_{data}}\sum_{i=1}^{N_{data}}\left(u\left(x_i, t_i\right) - u_i\right)^2 \tag{6}$$

Which ensures agreement with observed data and

$$L_{PDE} = \frac{1}{N_{PDE}}\sum_{j=1}^{N_{PDE}}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - v\frac{\partial^2 u}{\partial x^2}\right)^2 \tag{7}$$

enforces the equation's physics, where the sample points $\left(x_i, t_i\right)$ represent the initial location and time while $\left(x_j, t_j\right)$ covering the entire domain; W_{data} and W_{PDE} are the weights used to balance the interplay between the two loss terms [33]. These weights play an important role in improving the trainability of PINNs. The network is then trained by minimizing the loss via gradient-based optimizers until the loss is smaller than a threshold (ε).

However, not all domain scientists are convinced that AI can completely replace the NWP in the short term. In the near future, it is expected that people will continuously focus on a hybrid environment where numerical and AI models coexist for a long time. Further studies and directions are expected from the improvement of explainable and physics-based AI to enhance the model's trustworthiness and using advanced AI approaches such as self-supervised learning and transfer learning to improve the models' generalization capability. For instance, [37] from NVIDIA delivers AI model for weather forecasting.

2.5 Long-term prediction

Long term is a relative concept in Earth sciences and could have different durations within different domains. In geology, it could mean several thousands to millions of years for global and regional tectonism, while in meteorology, it is several months to years. The phrase "Long Term" is usually used during strategic planning and global-scale trends are required as supporting information. One typical example is to forecast climate changes at global level for next century [38]. However, based on the experimental results thus far, both AI and numerical models are struggling with long term predictions. The advantage of AI is not obvious over physics-based models when the forecasting time scale increases. This finding is understandable as the performance of AI depends on the quality of training data. For long term predictions, the training data coverage will become relatively insufficient and the quality decreases. Making AI learn long-term patterns is challenging, similar to the problems blocking the numerical modeling communities in the past decades. Thus, we think that for long-term forecasting, the speed of AI adoption will not be as fast as AI adoption for short-term prediction. We look forward to new revolutionizing techniques that could learn solid long-term patterns from limited training data and make accurate assertions about large-scale trends of Earth systems at a bigger time scale. Any progress in AI for long term prediction will have invaluable impacts to guide us in mitigating climate change and other grand issues facing our society.

2.6 Answering present questions using historical data and knowledge

In many scientists' impressions, AI is pictured as a robot which can answer any questions and provide instant appropriate advice based on history and context scenarios. How similar is this to reality? Recent natural language processing research has produced some eye-opening services that can deliver efficient question-answering performances on replying to chats, searching queries, following guidance, and finding quote sources. It is expected to see more and more research on training natural language processing models to digest the Earth science papers to answer relevant questions. Similar to other AI models, question–answer models also require high quality training datasets. There are some ongoing efforts to prepare science questions and answers [39]. More AI-ready science Q&A datasets are expected to be created in the near future, and intelligent answering services for Earth and environmental scientific questions are provided in the foreseeable future. Earth system sciences are sophisticated and contain much knowledge accumulated in many years of research and field work. It is important to make sure we have a strong workforce who have access to and understand this field of research. An AI model who can instantly answer the next generation's scientific questions and provide personal training will be critical to accelerate the research progress of Earth sciences.

Recent advancements in natural language processing have introduced transformer-based AI models with a self-attention mechanism. These models are revolutionizing how we extract knowledge from existing data. The self-attention mechanism acts as a dynamic spotlight, enabling AI to understand the complex interplay of scientific concepts within the text. As described in [40], self-attention calculates a weighted sum of values (V) using the similarities between a query vector (Q) and a set of key vectors (K). This mechanism helps in computing the context-aware representation of each word in the sentence considering all the other words in the sentence. This process can be mathematically described in the following equation:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}}) * V \tag{8}$$

Here, Q, K, and V are matrices that represent the queries, keys, and values, respectively. Each i-th row of the Q, K, and

V matrices corresponds to the i-th word in the sentence, and each j-th column of the matrices corresponds to the j-th dimension in the representation. $\frac{1}{\sqrt{d_k}}$ denotes the dimension of the key vector and acts as a scaling factor. The dot product of the query vector of the i-th word and the key vector of the j-th word is divided by the square root of $\frac{1}{\sqrt{d_k}}$ to ensure that the dot product values are within a small range of magnitudes. Then, the softmax function normalizes these dot products across all the words in the sentence to generate a probability distribution over those words. Then, the values of each word in the sentence are multiplied by the corresponding probability and summed to obtain the context-aware representation of that word. Thus, self-attention allows the model to capture the importance of each word in relation to the other words in the sentence, making it an effective tool for various NLP tasks, including reading comprehension, abstractive summarization, and textual entailment.

2.7 Exploration of unknowns

Many Earth scientists are wondering if AI can solve daunting science questions. Due to the fact that AI heavily relies on the training data and the patterns hidden in the historical data, many scientists doubt AI can find new knowledge outside the traditional unknown physics. It is true that most current popular AI models are probabilistic fitting and statistical machines, instead of intelligent reasoning engines. One key capability for exploring new knowledge is self-learning and evolving, which is a capability most current AI models do not have. We expect more intelligent AI models being proposed and tested to explore the unknown territory in Earth sciences, and provide real intelligence to reveal novel knowledge that was not discovered before.

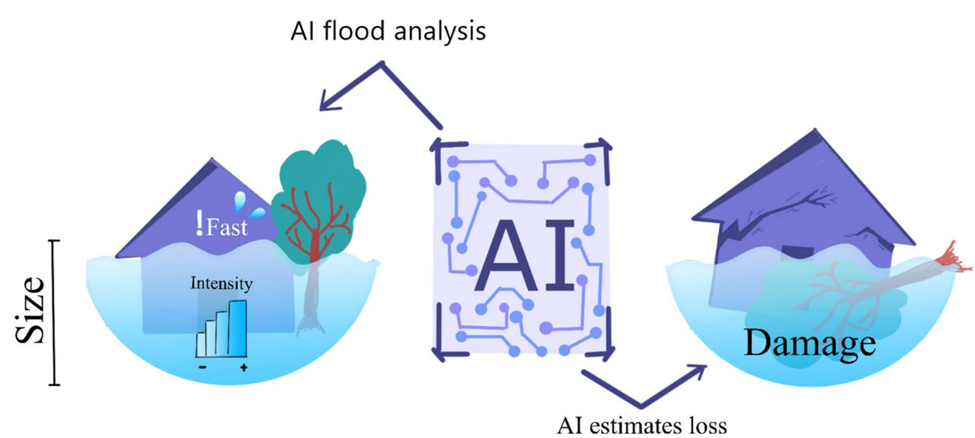
Some promising research at Columbia University has used AI to observe physical phenomena and uncover relevant variables which stimulate unexpected scientific discovery [41]. It could be considered a good start to use AI to extend our knowledge base. However, it still takes much effort to build general artificial intelligence to make AI models evolvable and less dependent on the scale and volume of the training data so AI models can do deductive reasoning based on the existing knowledge and derive new knowledge rules after combining with new observations.

2.8 Social impacts and applications

The stakeholders of geospatial datasets also include social scientists. Many social studies use remote sensing and stationary observations provided by federal agencies to analyze the relationships between social dynamics and environmental and climate changes [42–44]. Recent COVID-19 studies also use AI and geospatial satellite data to detect critical areas where individuals have a higher risk of contracting COVID (Atek, 2022). As shown in Fig. 7, AI methods have been used in natural hazard risk assessment, such as using AI for assessing the properties of the physical hazard itself in the flooding models or estimating the loss of system function given hazard loading [45, 46]. More applications of AI in the fields between social science and geospatial data sciences will be created to deliver more usable and actionable results to inform and guide our communities and country to better daily decisions and policy making.

In addition, ethical questions have been raised with the use of AI in Earth monitoring and predictive systems [47]. Social scientists can study and learn about the impacts of such technological evolvement and echo the concerns of our society on AI adoption and navigate Earth AI researchers to develop community-friendly AI services. Recently social science can play an important role in explaining the AI-derived products and their social impacts during interaction with real people. Social science researchers employ certain cognitive biases and social expectations to explain the AI process [48]. Similar to philosophy, cognitive and social psychology, AI development also needs to answer questions such as what constitutes an explanation, what the function and structure of the explanation is, and how to generate explanations and evaluate the explanations'

Fig. 7 AI for flooding response



quality [49]. Consequently, practical AI can build on existing research in social science since they provide a foundation for how people define, generate, select, evaluate and present explanations.

2.9 Data discovery and data curation

AI has been applied to assist researchers in quickly and accurately discovering Earth and planet science data they need. Generally, there are a variety of users of Earth science data, with varying levels of expertise and backgrounds. By leveraging the advances in AI fields such as NLP, data providers and distributors can help users find more relevant data. For example, NASA's distributed active archive centers hold EOS mission data and maintain seamless access to the data for users. For any DAAC to fulfill its mission, it is therefore essential that it be able to function as an effective data discovery tool. ML can be used by search engines to determine the most relevant results through an enhanced understanding of user search queries. Traditionally, search methods rely on matching explicit user search queries with indexed metadata. When the search query doesn't exactly match the metadata, a large number of searches can be missing. Modern NLP methodologies have recently been utilized to match queries with data through similarity metrics, as opposed to exact matches [50]. A similar, but different methodology, DAACs examine previous publications or applications of the data to offer users better datasets using NLP and graph models [51, 52].

2.10 Accelerating traditional models

One of the biggest challenges in adopting AI in the Earth science community is awareness of the power and pitfalls of AI. A few of these benefits and challenges are detailed here. Staying abreast of this fast-changing technology is difficult. The speed of data mining is one of the fundamental pillars for a functional modern society. The speed of data processing and information extraction and delivery is sometimes prioritized over accuracy and quality. The balance between speed and quality has been discussed for a long time in natural hazard response activities or other time-sensitive application scenarios. Near-real-time raw data products are available. For AI applications, the speed of the pipeline has several bottlenecks like the slow turnaround in data ingesting, model training, model prediction, and post-processing AI results into data products. The current solution focuses on the model training aspect, and the traditional way of speeding up computation like extract-transform-load parallel computing on powerful computing devices still struggles with the huge amount of the Earth science community datasets. The AI pipeline has similar time costs to numerical models, where retrieving data and preprocessing the data is time-intensive. Data pipeline engineering is an important component to speed up the AI workflow.

3 Best practices for implementing practical AI in Earth system sciences

There are many obstacles to overcome when creating usable AI models. Besides handling commonly known challenges [1] such as shortage of training samples, poor generalization, and lack of explainability, this section will focus on two more realistic problems for beginner AI scientists: how to use data and AI in the cloud, and community-oriented AI deployment and operation.

3.1 Project-specific AI product development and collaboration

Although the dream of general AI is being deeply exploited right now, most AI models still need to be carefully tailored for specific projects and certain well-defined tasks. There are basically two essential steps during the AI prototyping stage: problem definition and model development. Based on our experiences, finding correct scientific questions and giving a clean definition of the target AI tasks, is equally hard as actually developing AI models. It requires people with AI project experiences to help identify what problem AI is suitable and can help, and which ones it cannot. Generally, AI requires the presence of patterns in the dataset, meaning the dataset must not be random or close to randomness. The patterns don’t have to be completely explicit or instantly aware in human eyes, but should exist. The experiment part is generally standard protocol right now for AI projects. People gather datasets, especially datasets including ground truth or the training labels (most current AI tasks are supervised learning). Training data preparation is by far the most time consuming and needs the majority of the attention. It is not a sequential industrial pipeline-like practice. The data preparation and the model tuning are always done back and forth in many iterations. For example, people working on precipitation forecasting, may find some features like pressure, temperature, terrain, and land cover, are more useful in certain models and less useful in others. Therefore they have to prepare multiple training datasets to feed into different models to boost their forecasting capacity. In many projects, researchers have to manually repeat the iteration between data preparation and model tryout many times. Without proper project management, best practices like experiment transcripting and result seamless sharing, any AI projects can quickly collapse into a black hole which absorbs all the resources. These failed projects won’t deliver any good AI models, and have wasted numerous hours of researchers’ time and computational resources. That is also one of the main drivers of this paper to promote best practice on Earth AI research productivity by enforcing AI experiment recording and results sharing among team members or the entire community in a plain format that everyone can interpret. Current efforts like Geoweaver [53] have made a lot of progress on that and we need more efforts to dedicate to this aspect in ongoing and future Earth AI projects.

A great idea for an Earth AI project will definitely need more collaboration with many parties to help with important aspects such as technical support, funding support, computing support, usability support, user feedback, and potential market analysis and planning. No single person can achieve all of these aspects. Collaboration is one of the major requirements for most AI endeavors. Some important collaboration modes are the public–private partnerships, and research-industry-government collaboration. Government agencies like NOAA, NASA, NSF have already actively put out calls for AI products development to achieve various strategic goals in Earth sciences, and providing funding opportunities to connect resources with people equipped with knowledge, requirements, skills, hardware, to get things done. For the future generation workforce, NSF also has funding for universities to create programs to training students with AI/ML techniques to further discover the patterns in the geospatial datasets, and develop useful AI products to solve the challenging scientific problems, like earthquake forecasting, long-term meteorology forecasting, climate change and consequence prediction, and food security.

3.2 Community-wide AI deployment and production

As AI is gradually reshaping the landscape of geospatial data sciences, many research tasks are shifting from empirical manual analysis to data-driven automatic analysis. Earth science scenarios for AI models exist to accommodate end customers. How to deliver production-level stability and reliability is the biggest question for AI technologies. Most AI endeavors are struggling to meet the requirements of real world applications. Many tasks are not ready for deployment in AI. Scenarios like seismic signal explanation, hurricane forecasting, weather prediction, air quality simulation, and water discharge forecasting all need AI models to deliver not just accuracy, but also fast, explainable, reliable, and trustable results. That requires interactions between community users and AI models. For research users, they may directly get the AI models and deploy into their environment on lab servers or cloud platforms. For the public users of the product, data product teams are required to translate AI results into understandable format like maps or textual statements, such as, "There will be flash flooding in Fairfax County from 5:00PM to 6:00PM, please find shelter and avoid going out." which will require coordination with science communicators and public health specialists (Fig. 8). The community needs to transition the current information pipeline to knit AI models into the workflow. The user end portal likely needs to be changed as well to let people better understand the results, such as attaching a probability score with each prediction and linking the results to the provenance so geoscientists/meteorologists can verify and explain why such prediction is made (Fig. 8).

3.3 Maintenance & operation team guidance

The popular practice of operationalizing AI is MLOps (Machine Learning Operations), which usually refers to applying principles from the DevOps (Development and Operations) practices to the deployment of machine learning systems and includes monitoring the system to ensure it continues to work in the real world [54]. MLOps considers

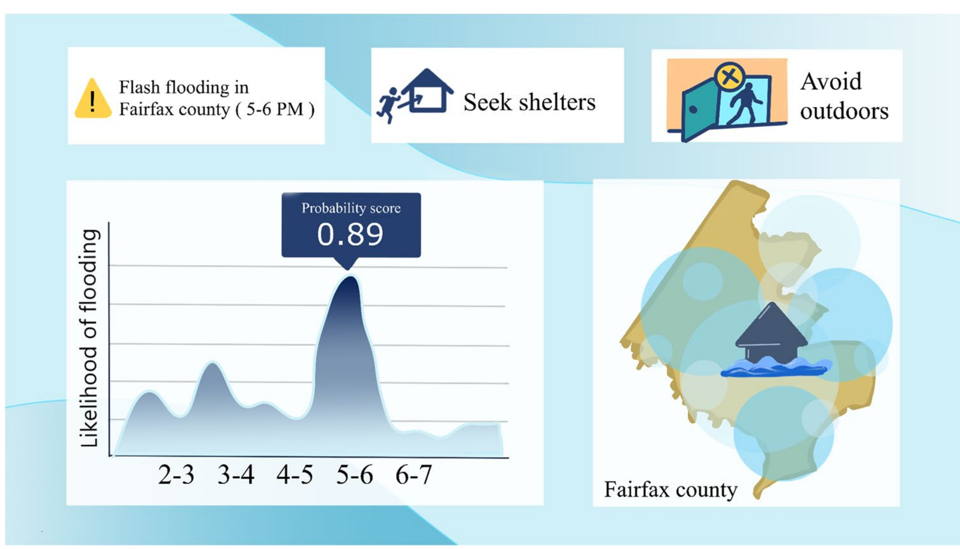


Fig. 8 User interface with AI for flood alert

the entire lifecycle of the model from data intake to the final use of the model. MLOps begins with exploratory data analysis, including understanding data quality and identifying particular issues. Model training occurs in the middle of MLOps, after the data has been cleaned and bad data has been removed. Between training and deployment, it is necessary to check model performance to ensure the model does not make any systemic mistakes. If the risks associated with misclassifications carry different real-world consequences, the model must be adjusted to maximize utility. For instance, in 2022, the city of Toronto deployed an AI model to predict whether bacterial levels at its beaches would be above or below the safety threshold [55]. If the consequences of beachgoers using an unsafe beach is worse than the consequences of a safe beach going unused, the model should err on the side of predicting "unsafe." During deployment and operations, the model must be monitored to ensure there is no decrease in performance due "data drift" (changes in the underlying data distribution) or other issues. With feedback from users and testing, the model can be adjusted to improve utility or at least avoid a decrease.

NASA's Interagency Implementation and Advanced Concepts Team (IMPACT) made MLOps for Earth observation a major component of its open source SpaceML Initiative [56]. The resulting MLOps tools reached Technology Readiness Level 9, ready for deployment, and can be used for satellites directed toward Earth, such as Worldview, as well as sky-oriented satellites, such as Hubble. MLOps is particularly important given that events of interest are often rare compared to the gigabytes to terabytes of data that are not of interest. SpaceML worked with high school students from around the world to provide cost-efficient data labeling [28].

3.4 AI auditing & accountability

Unlike the private AI projects, federal agencies' AI projects must comply with the Executive Order, "Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government," issued in December 2020 (Exec. Order No. 13960, 2020). In October 2022, the White Office of Science and Technology released a "Blueprint for an AI Bill of Rights," identifying five pillars: Safety and Effectiveness; Algorithmic Discrimination Protections; Data Privacy; Notice and Explanation; Human Alternatives, Consideration, and Fallback [57]. Additionally, the Government Accountability Office released a framework for AI accountability (GAO, 2021) to guide the auditing of AI applications for both federal agencies and other entities. It has four principles including governance, data, monitoring, and performance.

For geospatial data, the privacy risks may be limited, such as the use of low-resolution satellite data, or severe, in the case of phone and social media location data. It is relatively easy to deanonymize location data while allowing data users to track the movements of individuals to doctor's offices and other sensitive locations (Valentino-DeVries, 2018). Before engaging in an AI project using geospatial data, it is important to consider how the data and results can be used, or potentially misused. Notice is also important for anyone included in the data or impacted by the results. Developers should also provide a mechanism for people who are impacted by the applications or whose data are used in the development process to report the impact and withdraw their data from being used.

As more complex AI algorithms are developed, explanations of AI applications can be challenging to achieve. Explanations are particularly difficult for convolutional neural networks, which are commonly used for remote sensing data. A common solution is to use heatmap activations to ensure that the AI model is trained on the correct features of satellite data or other geospatial sources [58]. In all cases, it is necessary to not only build an AI model but to ensure it works as intended and does not produce negative unintended consequences. Some research suggests choosing inherently interpretable AI models instead of providing post hoc explanations using explainable AI (XAI) techniques, especially for high stake applications [59].

The Government Accountability Office's framework for AI accountability highlights the need to establish the process to manage, operate, and implement AI applications which ensure accountability. The process should be established at both the organizational level and AI system level. The organizational governance process allows the entity to engage with diverse stakeholders to ensure accountability and implement a risk-management plan. The system-level governance process provides technical specifications and procedures to continuously monitor the performance of AI systems at both component and system level and ensure that AI systems that are operated for intended uses [60].

3.5 Effectively using existing datasets and cloud computing

The effective utilization of geospatial datasets from institutions like NASA, NOAA, USGS, EPA, and public repositories hosted on data centers or cloud platforms such as AWS, Azure, and Google Earth Engine is redefining the landscape of data-driven Earth Sciences. These repositories offer an extensive array of invaluable Earth observations, including satellite imagery, climatic records, and environmental parameters. Accessible through robust APIs, they empower researchers to conduct intricate analyses and foster the development of AI models for comprehensive Earth systems' understanding. The scalability and computing power offered by major cloud platforms have revolutionized the capabilities of AI in Earth Sciences. With evolving AutoML solutions integrated into AWS, Azure, and Google Cloud, scientists, including non-programmers, can navigate complex AI processes through user-friendly interfaces, facilitating activities such as data preparation, model tuning, and deployment. These fully-managed AI services within cloud environments pave the way for swift prototyping and the development of practical AI solutions for Earth Sciences, underscoring the importance of aligning platform capabilities with research needs and considering cost efficiency and data quality in leveraging cloud resources.

Utilizing APIs like NASA's EarthData and Python libraries (e.g. earthaccess), researchers gain streamlined access to an extensive array of geospatial datasets, empowering them to conduct advanced analyses in Earth Sciences. The integration of Python with EarthData APIs offers a flexible and powerful means for querying and retrieving various datasets, streamlining the process for researchers to access specific data relevant to their Earth Sciences research. NASA, NOAA, and USGS have been actively involved in efforts to make their datasets cloud-native. Cloud-native format initiatives, such as NASA's adoption of cloud-optimized data formats like COGs (Cloud Optimized Geotiffs) or Zarr [61], have transformed the storage and accessibility of large-scale Earth observation data. These formats enable optimized storage and direct access to specific subsets of data without the need for complete downloads, thereby significantly reducing data transfer and storage costs. Researchers can use Python libraries to directly query and access cloud-optimized datasets from these agencies, enabling the retrieval of specific subsets of data for analysis without downloading the entire dataset. By integrating these cloud-native data formats with Python-based tools, researchers can perform large-scale analyses, including machine learning, deep learning, and statistical modeling, on these extensive datasets, allowing for comprehensive insights into Earth systems, climate patterns, and environmental changes.

For Earth AI beginners, cloud computing is becoming an important tool [62]. The traditional isolated local computing environments have been gradually replaced by open and publicly-accessible cyberinfrastructure, especially in the form of cloud computing. Generating large enough training datasets is, in many Earth science domains, a very expensive process. Big name data providers like NASA have already moved most of their datasets into the cloud. Availability of large datasets in cloud environments prevents the need for downloading by individual researchers, leaving more time for actual research (e.g., [63]). Most steps inside the full-stack AI life cycles depend on the availability of cloud computing and its offered technologies. However, the steep learning curve is a big challenge for new cloud users. The mainstream commercial clouds such as Amazon Web Service and Google Cloud are rapidly building their AutoML solutions on top of their gigantic cloud infrastructure. AI scientists can do everything inside their cloud environment. For non-programmer scientists, the cloud providers are forming low-coding environments (e.g., AWS SageMaker, Azure Machine Learning Studio, and Google AutoML) to allow them to finish training data preparation, model tuning, and service deployment by clicking buttons on a series of guided Graphical User Interfaces (GUI). Such cloud-native AI services are often referred to as fully-managed AI services for scientists to quickly prototype and build usable AI services. It is a reasonable path to get a production-level AI application for AI beginners. However, beginners have to fully understand the platform to compare the capability with their needs and consider the cloud costs and the training data quality before investing too much computing time.

3.6 AI workflow product management

3.6.1 AI maturity/readiness level classification

It is exciting to imagine AI becoming a routine tool in Earth sciences, to study and solve a wide variety of problems. When discussing specific AI products (either models or data products), they can generally be categorized into various levels based on their maturity and readiness for practical use, from proof-of-concept products to solid ready-for-use products. Industries like unoccupied autonomous vehicles have specific and detailed classification about the AI at different levels. According to the ElementAI classification [64], these are basic stages in operationalizing AI in production environments (Table 2). Similar classification can also be found in [65] with more fine-grained stage divisions.

Product maturity categorization system is not specific to AI products. The technology readiness level [66] defined many years ago at NASA has been widely used. Today the specifications to evaluate NASA product readiness are very detailed and include many mission-oriented requirements such as resolution and the removal of artifacts [67]. Besides NASA, there are several similar frameworks regarding data maturity and product/technology readiness in other government agencies. These frameworks are usually following some top guidelines to simplify the efforts by users to comprehend and take use of the eventual delivered products. For example, atmospheric scientists most likely will require the data products to be available in a standardized format, using a standard set of units and variable definitions for consistency and directly digestible by their analysis.

It appears that although the AI maturity frameworks have a lot in common with these existing maturity frameworks, there are still some differences. AI maturity needs to consider the interaction with its end users in daily usage, including all the components involved including model API, client software, computing platforms, and algorithm robustness in extreme conditions. The inherent uncertainty in AI models requires more detailed rules to regulate what kinds of AI

Table 2 AI Product Maturity Level [64]

Product Level	Definition
1—Planning	Problem identification based on the existing technology and available datasets
2—Experiment	Prototyping and experimenting with fractional datasets and wide variety of AI models
3—Development	Develop production-level data processing pipelines to apply the chosen AI model (with the best overall performance/cost rate) based on the originally collected training datasets
4—Production	Deploy AI into production for daily usage to digest streaming observational data and make in-time prediction. Also need to add other supporting functionality on top of AI like security and user management to interact with real users
5—Sustaining	Maintain AI services by iterating from servicing, problem feedback, retraining, to redeployment. Also with development in software and hardware, AI products should have upgrade plans for evolving to new models, algorithms, and technologies in future

products should be considered as “production ready”. An AI model might work well on the training data, but perform worse on new collected datasets, which is usually uncommon in the conventional non-AI technologies. A more relevant classification framework is the NASA technology readiness level framework [4], which could be appropriate to measure AI application’s maturity here.

3.6.2 End to end processing pipeline

The progress of AI research and adoption in Earth science domains is relatively slow. One main reason is lack of open and comprehensive end-to-end pipelines for AI workflows. The ideally expected deliverable for most Earth AI projects should be full-stack end-to-end workflows (or pipelines, can be used exchangeably), which are fully described and contain all the code files. It has been known that AI has a reproducibility crisis [68], due to the randomness and black-box nature of the involved models. Figure 9 shows an example workflow replicating the Kenya crop mask mapping workflow [69]. Each circle represents a Python process, and the connections indicate the order of execution. Most processes are actually for data preparation. Machine learning model training and testing account for only about a third of the processes. Unfortunately, most people only share the ML portion of processes and give less attention to the other processes, especially those essential upstream steps.

ML researchers look for solutions to better compose and share the E2E ML workflows, and one of the popular efforts is trying to improve geospatial AI FAIRness. FAIR principles are not limited to dataset but also the other project objects such as tools workflow [70, 71]. So far, the reality of the current AI for Earth Science still deviates from these principles, therefore, hindering the reproducibility of AI. The challenge of reproducibility may come from various aspects, for instance, the changes in the systems, software versions, the nature of the training ML models involving randomness, and the ML frameworks using various precision to accelerate the training procedure. Figure 9 shows an example end to end workflow using NASA remote sensing data to map cropland in Kenya.

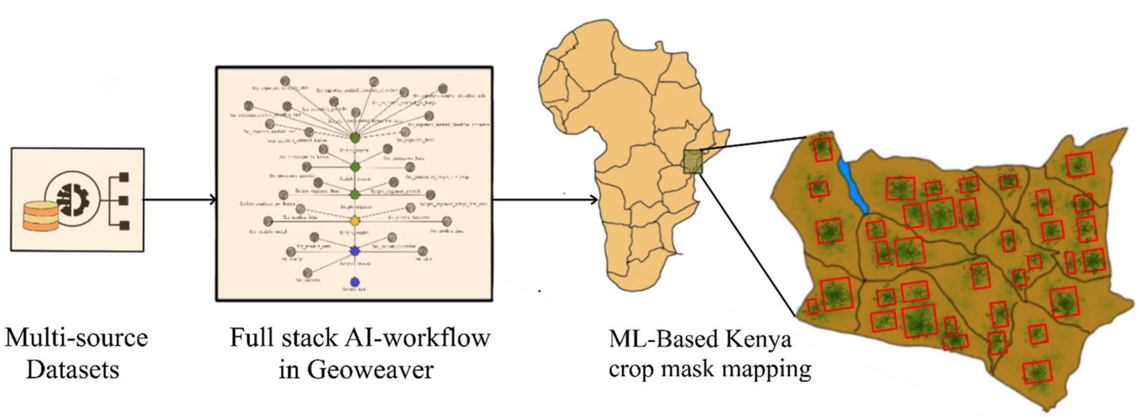


Fig. 9 Example AI workflow of ML-based Kenya crop mask mapping in Geoweaver (https://github.com/earth-artificial-intelligence/kenya-crop-mask-geoweaver)

4 Example success use cases

As mentioned above, it is very challenging to realize "one-size-fits-all" AI products, given the current AI techniques it is hard to implement "wide", general purpose AI that can be used across multiple problems or domains. Instead, most current AI models are trained for a specific "narrow" purpose, with a similarly narrow, specially prepared training dataset. There is no straightforward rule to determine when an AI model is considered good enough for production. Researchers need to compromise to determine which model is ready to be delivered to its intended end users. Generally, if a model can work with an inside-boundary accuracy for a specific purpose, it is considered good enough. We call it an "fit-for-purpose" model which is not the best possible model but it can do the work. For example, for snowfall forecasting, suppose model A outputs 91% accuracy while model B gives 89% overall accuracy, while model A costs two times longer than model B. For most users, model B would be the on-purpose model as its overall performance combining accuracy and costs is the best. To give a more concrete understanding, the following sections will briefly introduce some on-going efforts within the geospatial data science community to make AI practically usable for their users.

4.1 Ozone forecasting

The CMAQ model, also known as the Community Multiscale Air Quality model [72], is widely utilized by atmospheric scientists to predict changes in air quality. It measures various parameters such as ozone, PM2.5, NO2, SO2, among others. However, CMAQ has consistently exhibited overestimations and underestimations in specific U.S. regions. To address this issue, the CMAQ AI team at George Mason University conducted a feasibility analysis to leverage machine learning for calibrating CMAQ results. Subsequently, they implemented this approach in an experiment. Utilizing Geoweaver [73], a workflow was composed for proof of concept, and the results from training a random forest model on the 2021 dataset were highly promising. When evaluating the ML-calibrated ozone results, it became evident that they consistently outperformed the original CMAQ results. As a result, the GMU team seamlessly integrated this workflow into the production environment, allowing for the generation of daily ozone maps. The small circles within the maps represent ground truth data collected by the AirNow station network [74]. Figure 10 is an illustration of the comparison between AI and CMAQ results.

4.2 Underwater image recognition

NOAA AI strategic plan [76] projects that AI methods are expected to boost transformative advancements in the quality and timeliness of atmospheric science, products, and services. One of their preliminary efforts is using ML in detecting organisms in the captured underwater images. Currently, underwater surveys within NOAA fisheries require a large amount of manual oversight by data analysts to interpret the images. That is not sustainable as the amount of images is rapidly increasing. For example, the NEFSC (Northeast Fisheries Science Center) Habcam (Habitat Mapping Camera) benthic survey [77] now collects approximately five million images a year. AI practitioners turn to AI for help, and developed VIAME (Video and Image Analytics for a Marine Environment) convolutional neural network [78], which has been tested and proven very promising at automating identification of the organisms and relieving the human analysts

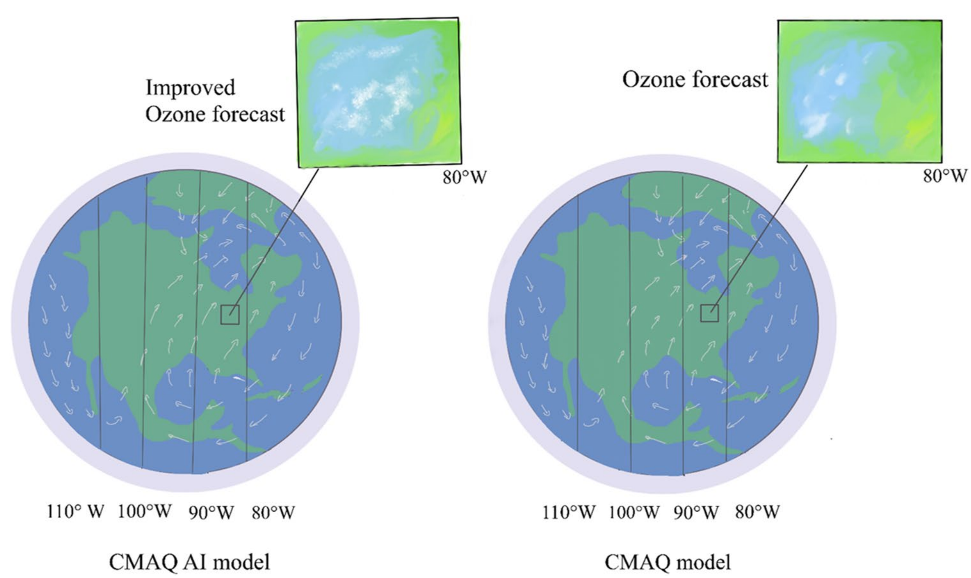


Fig. 10 Comparison of AI and physics model results [75]

from the heavy burden of manually reading the huge number of images (Fig. 11).

4.3 Land cover map downscaling

Within the Earth science community, a prominent and pressing demand revolves around the enhancement of current data products, primarily addressing the limitation of coarse resolutions that hinder the extraction of actionable information. A breakthrough success story in addressing this challenge emerges through the innovative use of AI to refine images, providing unprecedented clarity and intricate details. A prime illustration of this triumph is found in the NOAA NOS (National Ocean Service) C-CAP (Coastal Change Analysis Program) program's initiative. Through the strategic implementation of AI, this program has achieved a remarkable feat by downscaling land cover maps designed for U.S. coastal regions. The advancement is staggering, transforming the resolution from a relatively coarse 30 m to an impressive 1-m resolution [80].

The magnitude of this achievement cannot be overstated. The significance lies not only in the numerical leap from 30 to 1 m but, more importantly, in the tangible impact on map users. The enhanced resolution enables users to discern intricate details and objects with unprecedented clarity. What was once invisible or excessively coarse, hindering meaningful interpretation, is now accessible and distinguishable. This success case serves as a testament to the transformative power of AI in addressing critical challenges within the Earth science domain. It showcases not only the potential for technological advancement but, crucially, the tangible benefits realized in providing Earth scientists with clearer, more detailed information that was previously elusive.

4.4 Coral reef detection

Similarly, Coral detection is another headache problem for survey image interpretation which was conducted manually by analysts at a low efficiency. Scientists developed CoralNet [81], which can annotate coral reef images and automatically distinguish different species of corals, and has greatly assisted scientists by saving them a big amount of time on labeling. The latest version of CoralNet can also provide higher resolution products and includes script-level access to allow interfaces with other projects [82]. The model providers have benchmarked a bunch of machine learning models before settling on the EfficientNet to train the official version of the CoralNet model and deploy them into use after successful beta trials.

These examples showcase how AI works in real life to address specific Earth scientific problems which are very cumbersome for the existing approaches and used to involve heavy manual human supervision. They also prove that AI can do things impossible before like providing data in greater resolution due to automated data infusion. It should be noticed that due to the specialization of AI models, each use case is different in terms of daily operations. For example, CMAQ forecasting's model inputs are continuous and time sensitive so it needs to be run every day on a regular basis, while the CoralNet is only triggered when users request new coral images. Also, it is obvious that no AI models are perfect and when errors happen, the operation team needs to respond and fix the issues in a timely manner, which echoes the "fall back" guideline by the White House AI Bill of Rights.

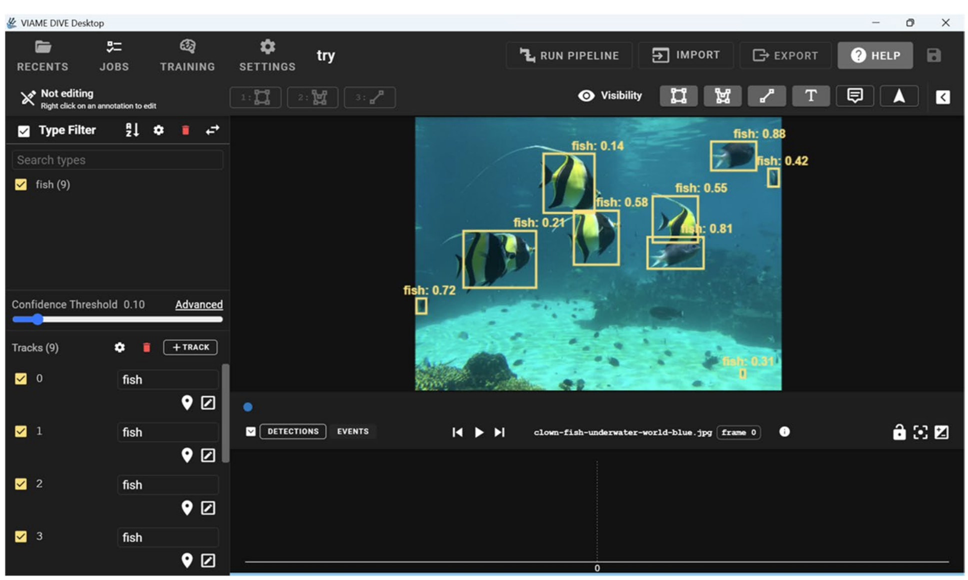


Fig. 11 VIAME underwater recognition (the number after fish is AI confidence score about the label) [79]

4.5 AI in volcano science

Understanding and predicting volcanic eruptions is an important subset of the geosciences, however volcanic systems are vastly complex, and each is unique. These complicated systems still leave volcanologists with many questions: when will this volcano erupt next? Why will it erupt? How large of a magnitude will the eruption be? And more importantly how long will it erupt? Machine learning offers new avenues to explore these questions and to utilize multidisciplinary datasets. Presently, machine learning has been utilized to improve the quality of magmatic pressure and temperature estimates, allowing us to visualize the spatiotemporal dynamics acting in volcanic plumbing systems. Volcanic minerals (e.g. clinopyroxene; amphibole; olivine) are compositionally pressure and temperature sensitive, meaning at different pressure and temperature conditions the elemental composition systematically shifts. Thus, looking at the changes in chemistry within different regions of a mineral allows us to track the magmatic history of a volcano prior to the eruption. Recent developments turned to using machine learning random forest methods to tackle this problem and have improved the accuracy and precision of these estimates [83, 84]. These methods utilize large open-source datasets of experimentally derived minerals, which crystallized at a known pressure and temperatures, to train the model. Improvements to the error of these models allows us to see the depths of the earth in better resolution, and thus aid us in understanding processes that trigger volcanic eruptions and in detecting signs in deep magma dynamic variations that may be reflected through changes in eruptive behavior (Fig. 12), for example, the volcano monitoring platform, MOUNTS [85]. Other machine learning approaches in volcanology are also being explored to classify and interpret seismicity related to volcanic activity, which often come with large datasets and are important for real time volcanic monitoring [86–88]. Machine learning might even be the way forward to increase our ability to predict volcanic eruptions. For example, Ardid et al., [89] utilizes cross correlation machine learning to generate a predictive forecasting model, which can improve short term eruption warning systems for gas driven volcanic eruptions. Machine learning advances will drive the field forward and aid us in answering our fundamental questions about volcanoes and allow us to better mitigate and respond to these hazards.

5 Challenges of using AI in the geosciences

This section examines the general need for artificial intelligence (AI) from the perspective of data scientists and data users in the Earth science community. Based on the current datasets, the research community is actively exploring AI models to generate higher quality socioeconomic products that are directly relevant to decision making in our society. The community is exploring AI to address the problems that are currently nearly impossible to solve by traditional research approaches. There are strong needs in data-driven sciences right now. This section will detail these needs.

5.1 Tackling missing data, biases and uncertainties

AI requires continuous and high-quality data for training and testing. The Earth science community has a tremendous amount of data available. However, there is still a dearth of high quality data concerning the key variables, accuracy, and spatial–temporal coverage. In the NASA community, remote sensing reflectance data is the major driver behind the big data archive and there are numerous thematic data products derived from reflectance using algorithms and models, e.g., land surface temperature, snow cover, land cover, precipitation, soil moisture, soil temperature, air temperature,

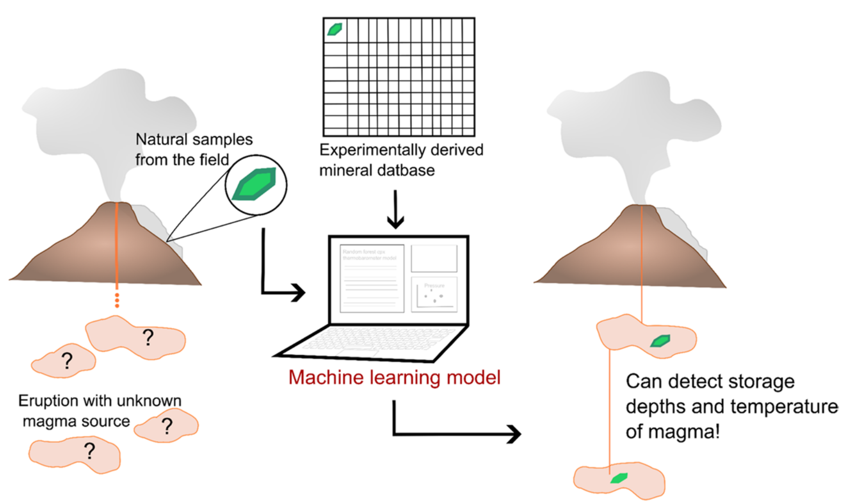


Fig. 12 Machine Learning model used to detect magma storage conditions

albedo, etc. However, currently only a fraction of these data products can be used as training labels because of the concerns over resolution and precision. Instead, NASA data is commonly used as input variables in the training dataset. Clouds and other weather conditions also caused long-term missing data of continuous observation of the land surface in remote sensing datasets (Fig. 13). Satellite SAR/LiDAR data can penetrate the clouds and operate day and night. However, the limited resolution, long revisiting period and high ratio noises in the signals degrade the advantages of satellite SAR/LiDAR data over optical satellites. These missing areas pose major problems when applying AI to digest and train on NASA datasets. The designers of feature engineering of Earth AI models must be knowledgeable about these satellite datasets and clearly understand the issues caused by missing data. They should also choose carefully what data should go through the training process. Data gaps are a general issue in geospatial datasets, and not limited to AI use cases. However, these gaps will be a critical issue since it results in incomplete or biased learning patterns of AI due to missing information. As a result, trained AI models can fail to recognize correct patterns, miss important signals, or even become unusable in real world scenarios.

Besides the data bias, algorithmic bias, or the tendency of AI models to amplify the data bias (or its inability to correct them) are another major concern. For locations which are poorly represented in the training data [90]. AI generally has lower accuracy and generalization. To be practical, AI systems should acknowledge users with these issues and the known limitations by delineating the scope of model applicability [91]. AI models should only be used for locations/regions at a certain time range with comprehensively-tested trustworthy reliability [92]. To produce scientifically useful products, AI practitioners need to understand and quantify the errors, biases, and uncertainties in the used data [93]. Methods like probability-based sampling can be used to improve the quality of ground truth data to train the AI models [94–96]. Also, more efforts should be made to assess the biases in the AI outputs at various spatial–temporal scales to track the potential error origins during data aggregation [97]. As for quantification of AI uncertainty, distribution-free methods are usually recommended [98].

5.2 Preparing AI-ready data

It is an unspoken truth that geospatial data scientists spend most of their time preparing rather than analyzing data [63]. The challenge with this process is that it does not only extend the ML cycle for experienced practitioners, but creates a high entry barrier for those with less experience. There has been little work done to develop a detailed understanding of the challenges associated with the data before it is used to build a model. Furthermore, geospatial data possess attributes that require special attention. According to the FAIR (short for Findable, Accessible, Interoperable, and Reusable) guiding principles [99], research data must be Findable, Accessible, Interoperable, and Reusable. Moreover, many AI/ML applications rely on inherently opaque model architectures [100]. Consequently, the quality and integrity of the data becomes even more vital for AI applications. However, data quality information is difficult to generate, curate and share, especially across disciplines [101]. Improvements have been made over the years such as the Data Product Development Guide (DPDG) for Data Producers. DPDG, data quality is

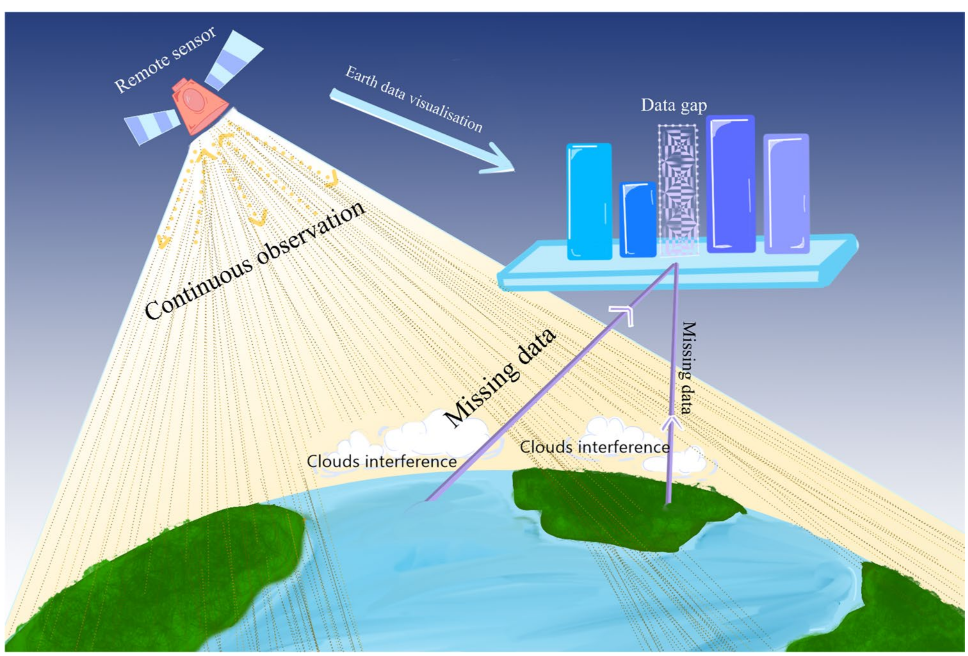


Fig. 13 Missing data illustrations

included in metadata, was developed as a suggested practice for NASA Earth Science Data Systems [102]. Accessing AI-ready data is another challenge and most data centers do not provide such services (e.g., extreme precipitation). For data to be found, it is essential that it be accompanied by standardized metadata that is understandable by search and automation tools. Community-accepted standards, such as the SpatioTemporal Asset Catalog (STAC), allow users to find data both spatially and temporally.

Consumers of ML data historically downloaded their required data into a local machine or an on-premise HPC system. This has turned out to be extremely inefficient and time consuming. At the same time, public and private cloud services are becoming increasingly more common and affordable. This provides the opportunity for data providers to store their data in cloud-friendly formats, such as Zarr and cloud-optimized Geotiffs (COG). Combined with streaming APIs that feed data directly into model training platforms, users can take advantage of the data they need without downloading large datasets. In order to make data more interoperable, it should be noted that different communities may use the same data in different formats. By offering data in formats commonly used across these communities, data providers can make data more interoperable. Alternatively, datasets could be accompanied by necessary preparation codes that could not only convert data formats but also prepare the data for downstream training frameworks (Tensorflow Dataset module or PyTorch DataLoaders).

A best practice is to develop benchmark datasets for certain domain problems and share with the entire community. Many benchmark datasets have been used in other domains and powered breakthrough AI projects. In comparison, the Earth science community has less available benchmark datasets with equal high quality. However, researchers have recently started to catch up; we see an increasing trend of efforts on creating benchmarks. Generally, a benchmark dataset should target a specific domain problem and be widely accepted as a common asset among the community. The data should be reusable and model independent so researchers can experiment various AI models on them and do intercomparison. In addition, to ensure the objectivity of the intercomparison, the benchmark datasets should be evaluated with the same set of evaluation metrics, a clear problem statement, and an unified way to read data using standard high-level language such as Python and R [103].

5.3 Reducing experiment & operation costs

Earth scientists are spending tremendous amounts of time on modeling efforts to understand how the Earth systems work and how they are going to change under various conditions such as climate changes and human–environment interaction. The current practice is to develop multiple models and run them in parallel to find out which prediction has the most consensus among those models. Every model runs and digests huge amounts of data and takes substantial computing resources which have to be done on big computing clusters or even supercomputers. Some of the numeric processes have been considered time consuming and a better approach to complete them is needed. There are also increasing concerns about the carbon footprint of these resource-intensive models [104]. Currently, the modeling community is turning to AI, which consumes relatively few resources and employs straightforward strategy to learn the underlying patterns of how input variables impact the target phenomenon. Introducing AI models into the existing numerical models will have the benefit of supplementing those missing links to replace the problematic and computation intensive processes in the numerical models, and reduce the overall costs of the modeling efforts (as shown in Fig. 14).

On the other hand, AI engineering has a high initial cost but brings more benefits and accelerates the modeling efforts in a long-term perspective, which will save huge amounts of resources and costs. If data scientists want to reuse the existing AI models, the process is easier because AI supports transfer learning and standardized pipelines of how the model is relinked and reused. Tuning efforts of AI models is also less time-intensive than tuning numerical models. Hyperparameter tuning has already been automated at some level by techniques like parameter searching cross validation (e.g., GridSearchCV or RandomizedSearchCV). The burden on the modelers or data scientists will be much smaller and they will save big chunks of time to focus on the Earth's scientific problems instead of modeling technical issues.

5.4 Realizing physics-based AI

Another ongoing interesting development is physics-based AI. The created models are often referred to as 'hybrid' models, combining the traditional process-based models with data-based approaches. Physics-based AI can provide more structure than a 'data-only' approach, and is a potential solution for biased datasets. As of yet, there is no clear consensus on *how* to develop physics-based AI solutions: these range from implementing AI into existing process-based models, or adding a set of (limited) physical constraints to a data-driven system. In the latter case, the physical constraints are often fundamental laws such as the conservation of mass, energy or momentum. The degree of physics inclusion is also often used to distinguish between different AI approaches. Recently, scientific progress has been made in the field of hydrology by using physics-based approaches, illustrating its potential when traditional approaches provide insufficient insight [105, 106].

A recent review paper [33] has pointed out that many current studies are trying to incorporate physical laws into

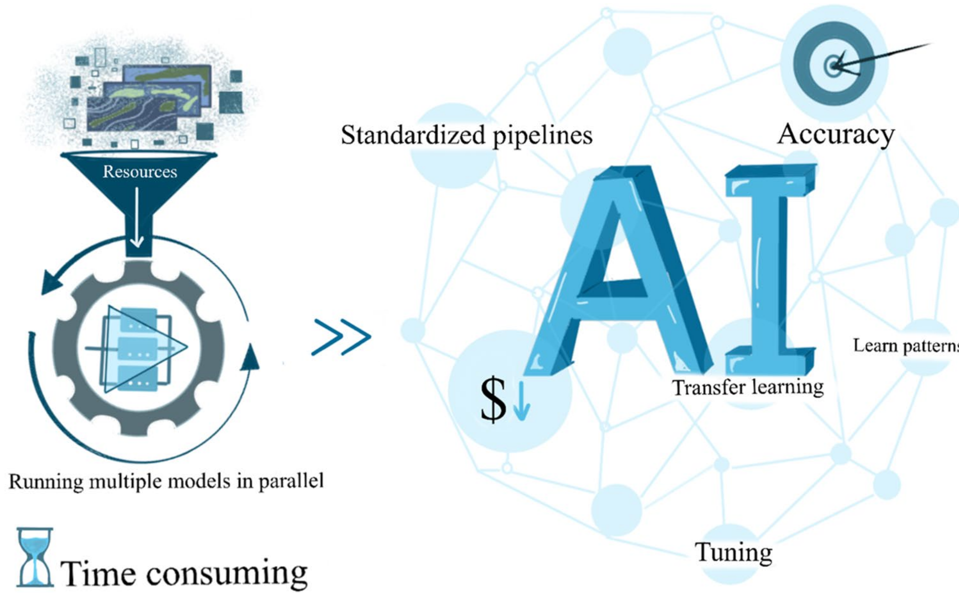


Fig. 14 AI advantages over existing models on costs and efficiency

machine learning models to achieve better accuracy, lower training costs and improved model applicability at spatio-temporal dimension. The Earth science community is actively seeking solutions to inject our already learnt knowledge into the AI methods to address challenges that are inherently suited to AI methods. A number of Python libraries are under development to help build physics-informed AI models for various science problems. For instance, Zhu et al. [107] used a novel parameterization approach to train a neural network with physical constraints incorporated on hydrographic and turbulence observations in tropical Pacific. It shows that adding physical constraints can improve the generalization of trained neural networks, but many issues like uncertainties and requirements for high quality training samples still remain. The observed data does not always obey physical laws, but the models driven simply by data are inaccurate as well due to some hard-to-control factors like the low-quality training samples. A natural idea to boost the performance is to incorporate the laws of physics and the data-driven AI models, with a wish to combine the advantages and achieve something better than using either way alone. The current research on physics-informed AI is at its infancy stage, but with great potential only limited by design imagination. This is a very promising direction that can accelerate our Earth AI development by using both methods, physics-driven methods and data-driven methods, allowing the modelers to 'walk with two legs,' as it were.

5.5 Making AI operational

AI research and development is still at the prototyping stage in the NASA Earth science community. "Being in operation" means making AI products accessible and usable by the general users without unexpected interruption and provided with customer support. Strategy and infrastructure to transfer AI research into Operation (R2O) are the required components to guide the transition from AI research team to the production team and communicate over issues, feature requests, customer support, etc. New engineering practical strategies like DevOps or recently the concept of MLOps [108], have emerged and become popular in the industry. MLOps denotes a collection of practice and skills to deploy and maintain machine learning models in production reliably and efficiently. It combines the software engineering cycles with ML model development and version tactics. Most importantly, the testing and validation in ML will be automatically conducted by continuous integration testing software such as Circle-CI, Github Actions, or Jenkins, like data validation, model improvement verification, model service integration, etc. Similar practices might be adopted to roll out the AI products in the NASA data science community. However, validating satellite-based products on a global scale often requires in-situ observations which can be difficult to acquire, especially in remote regions and over oceans. Validation is often continuous work [23]. From the perspective of data scientists, operational AI will provide them with a much more powerful capacity to understand the Earth via the enlarged enhanced lens of high-resolution continuous data on key variables.

6 More discussion points

6.1 Training data quality

The key factor in making AI more practical and usable is the availability of high-quality data. Scientists need access to large amounts of well-curated data to train AI algorithms and validate their results. Governments, universities, and private companies can play a role in making data more widely available to the scientific community. AI tools need to be easy to use and accessible to geoscientists. This requires developing user-friendly interfaces that allow scientists to easily input data, run analyses, and interpret results. Interfaces should be intuitive and should require minimal technical expertise. AI algorithms need to be robust, reliable, and accurate. This requires ongoing development and improvement of AI algorithms, as well as testing and validation using real-world data.

6.2 AI Practitioners are in the driver seat

“If I had asked people what they wanted, they would have said faster horses.” The famous Henry Ford “faster horse” quote [109] reflects that every big innovation and great novel product requires the people with first-hand experiences on the products to defend their views, make bold attempts, and bring what they see as the best future avenues into reality no matter how different they are from the paths that others without the firsthand experiences may promote. AI practitioners should actively reach out to the intended users and hear their requirements; however, AI practitioners must also envision and figure out how AI products can target core scientific problems without projects being waylaid by tedious technical details. Similar to how the iPhone touch screen replaced the traditional keyboard, AI will bring a lot of surprising changes to the existing research routine of Earth and environmental scientists. The time-consuming cumbersome work can be replaced with button-clicking effortless steps with the help of AI services.

6.3 Preparing next generation AI-ML practitioners

Current students will be the next generation's AI scientists. Therefore, it is essential for students to have data literacy skills for them to adequately apply AI techniques in respected fields of applications. "Data dexterity" is the skill or ability to gain the skills vital for the application and development of AI techniques. As every AI technique depends on the data that is being used and fed into the models, we must expect our students to demonstrate a higher level of data dexterity to apply AI techniques properly. Future AI practitioners need to demonstrate how to use strategic thinking combined with a solid technical foundation in topics such as data lifecycle, data management, workflows, metadata standards (e.g., the NetCDF Climate and Forecast Metadata Conventions [110]) etc., before implementing any AI based models. Munasinghe et.al [111] describes that the ability to properly utilize data to apply relevant analytical methods to solve and formulate science and engineering problems is also a subset of skills future and current AI practitioners require. Some pilot projects have been conducted, implementing data-intensive courses using NASA datasets in class projects and assignments to promote the data dexterity among students.

Nowadays, people can easily learn many advanced techniques purely based on the free open materials online rather than in college programs. Informal AI training can also contribute to the formation of the next generation of AI experts in geosciences. Many geoscientists do not have a degree in computer science or AI majors, and that should not be an issue. Hackweeks and mentorship training models are some existing modes of education and community engagement [112]. Hackweeks are time-bounded events that blend elements of a summer school with open project work or "hacking" [113], and are designed in a way that encourages immersive, interactive learning in a space that is welcoming to people of all backgrounds. The content within a hackweek is intentionally designed co-creatively with participants, and there are facilitated opportunities for networking and community building. As an example, the NSF-funded GeoScience MAchine learning Resources and Training (GeoSMART) project seeks to develop a dynamic and sustainable curriculum by integrating education, cyberinfrastructure, and research that can be also brought into the classroom (https://geo-smart.github.io/). The GeoSMART educational pathways build upon three levels characteristics in research: i) basics in data science and computing, ii) machine learning tools, iii) research project development and deployment of the learnt tools to large-scale data with cloud computing. The hands-on component of the GeoSMART framework is built around the hackweek model [113].

6.4 What role can non-AI techniques play in practical AI?

Practical AI includes a stack of technologies, of which AI models and algorithms are only a fraction. It can be expected that new technologies will keep evolving within the AI landscape, and embedded as part of the overall AI solutions. Since the rise of large mega-models, driven predominantly by deep neural networks, many applications have focused on leveraging these models to solve problems from start to finish. Nevertheless, as the artificial intelligence community exhausted many of these solutions by creating more complex models with more data, their underlying shortcomings began to emerge. Today, it is generally acknowledged that these models can provide a variety of task-level functionalities. However, these models are typically complemented by "traditional" and "non-AI" techniques in practice. The frontier research domains, such as knowledge graphs and model-based reinforcement learning, could be overlapped with AI and will have a great potential in practical AI. For example, opposed to the hard-to-explain AI models, knowledge graphs can help to see the patterns in explicit form of rules, and understand the completeness of the datasets and components of the model. New technologies can also help AI make sure the training data does not exclude specific extreme use cases, creating a distribution different from the real world, to improve AI generalization and inclusiveness. Overall, there is no exclusive definition for what technologies can be used in practical AI. On the contrary, there are many possibilities where AI can benefit from other non-traditional AI techniques.

7 Conclusion and outlook

AI is a very popular topic within the Earth science community and government agencies, and many groups are spending tremendous amounts of effort to make it practically usable in solving scientific problems. This paper captured some of the issues and future horizons for AI in the Earth science community. These issues can be used by future AI practitioners when planning out their research projects. Most scientists are wondering about the potential pros and cons of AI before seriously using AI in their research routines and operational scenarios. This position paper aims to picture the landscape of AI-involved data-driven applied sciences by discussing the current and upcoming needs of the research community, what practical AI looks like, how to realize practical AI in NASA and the broader research community based on the current techniques, and the expected outcome including both benefits and issues. This paper also discusses some further topics concerning the unavoidable changes in the near future such as the fast evolution of the AI foundation models and how the NASA community should adapt. This paper provides an important reference to the geospatial-data-driven science community to adjust their research road maps and allocate resources to make their AI work more practical in real world scenarios.

Acknowledgements Thanks to the support from all the colleagues, especially people from NASA ESDSWG and ESIP ML cluster. All the information and opinions in this paper are made by the authors, and do not represent the opinions of their affiliated institutions. Special thanks to Dr. Kelly Monteleone for reviewing the manuscript.

Funding This research was partially funded by NASA ACCESS (#80NSSC21M0028), NSF Geoinformatics program (EAR-1947893 & EAR-1947875, 2020), NASA Health and Air Quality project (#80NSSC21K0512, 2018), NSF Cybertraining (#2117834, 2021), and NOAA (#NA22OAR4590512).

Data availability This paper has no associated data.

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer The statements, findings, conclusions, and recommendations are those of the authors and do not reflect the views of NASA, NOAA, EPA, or any part of the United States Government.

References

1. Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S.M., Wang, J., Lin, C., Cristea, N., Tong, D., Carande, W.H., Ma, X.: others: A review of earth artificial intelligence. Comput. Geosci. **159**, 105034 (2022)
2. Sun, Z.: Actionable AI for climate and environment. In: Actionable Science of Global Environment Change: From Big Data to Practical Research. pp. 327–354. Springer (2023)
3. Sun, Z., Cristea, N.: Introduction of artificial intelligence in Earth sciences. In: Artificial Intelligence in Earth Science. pp. 1–15. Elsevier (2023)
4. National Aeronautics and Space Administration (NASA) Earth Science Technology Office (ESTO). Technology Readiness Levels (TRLs). (n.d.). Retrieved from https://esto.nasa.gov/trl/
5. Wang, J., Bretz, M., Dewan, M.A.A., Delavar, M.A.: Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. Sci. Total Environ. **822**, 153559 (2022)
6. Abdi, A.M.: Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience & Remote Sensing. **57**, 1–20 (2020)
7. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N.: Prabhat, fnm: Deep learning and process understanding for data-driven Earth system science. Nature **566**, 195–204 (2019)
8. Sun, Z., Cristea, N., Rivas, P. (eds.): Artificial intelligence in earth science: best practices and fundamental challenges. Elsevier (2023)
9. Koren, G., Ferrara, V., Timmins, M., Morrison, M.A.: Global Environmental Change Perspectives on Integrated, Coordinated, Open, and Networked (ICON) Science. Earth and Space Sci **9**, e2022EA002231 (2022)
10. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.-H.: Diffusion models: A comprehensive survey of methods and applications. ACM Comput. Surv. **56**, 1–39 (2023)

11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Commun. ACM **63**, 139–144 (2020)
12. Nassif, A.B., Talib, M.A., Nasir, Q., Dakalbab, F.M.: Machine learning for anomaly detection: A systematic review. Ieee Access. **9**, 78658–78700 (2021)
13. Bergen, K.J., Johnson, P.A., de Hoop, M.V., Beroza, G.C.: Machine learning for data-driven discovery in solid Earth geoscience. Science **363**, eaau0323 (2019)
14. Xiong, Y., Zuo, R.: Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. Comput. Geosci. **140**, 104484 (2020)
15. Yang, J., Xu, R., Qi, Z., Shi, Y.: Visual anomaly detection for images: A systematic survey. Procedia computer science. **199**, 471–478 (2022)
16. Schmidl, S., Wenig, P., Papenbrock, T.: Anomaly detection in time series: a comprehensive evaluation. Proceedings of the VLDB Endowment. **15**, 1779–1797 (2022)
17. Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., Tang, Y.: A brief overview of ChatGPT: The history, status quo and potential future development. IEEE/CAA Journal of Automatica Sinica. **10**, 1122–1136 (2023)
18. Elhafsi, A., Sinha, R., Agia, C., Schmerling, E., Nesnas, I.A., Pavone, M.: Semantic anomaly detection with large language models. Auton. Robot. **47**, 1035–1055 (2023)
19. Cao, Y., Xu, X., Sun, C., Huang, X., Shen, W.: Towards generic anomaly detection and understanding: large-scale visual-linguistic model (GPT-4V) takes the lead. arXiv preprint arXiv:2311.02782 (2023)
20. Tur, A.O., Dall'Asen, N., Beyan, C., Ricci, E.: Exploring diffusion models for unsupervised video anomaly detection. In: 2023 IEEE international conference on image processing (ICIP), pp. 2540–2544. IEEE (2023)
21. Guzzetti, F., Gariano, S.L., Peruccacci, S., Brunetti, M.T., Marchesini, I., Rossi, M., Melillo, M.: Geographical landslide early warning systems. Earth Sci. Rev. **200**, 102973 (2020)
22. Bortnik, J., Camporeale, E.: Ten ways to apply machine learning in the Earth and space sciences. In: AGU fall meeting abstracts, vol. 2021, p. IN12A-06 (2021)
23. Nguyen, P., Shearer, E.J., Ombadi, M., Gorooh, V.A., Hsu, K., Sorooshian, S., Logan, W.S., Ralph, M.: PERSIANN Dynamic Infrared-Rain rate model (PDIR) for high-resolution, real-time satellite precipitation estimation. Bull. Am. Meteor. Soc. **101**, E286–E302 (2020)
24. Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S.: Skilful precipitation nowcasting using deep generative models of radar. Nature **597**, 672–677 (2021)
25. National Oceanic and Atmospheric Administration (NOAA). NOAA and OSTP workshop on artificial intelligence and weather prediction (2024). Retrieved from https://www.nesdis.noaa.gov/news/noaa-and-ostp-workshop-artificial-intelligence-and-weather-prediction
26. National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory (PSL). AI for numerical weather prediction (AI4NWP) workshop (2023). Retrieved from https://psl.noaa.gov/events/2023/ai4nwp_workshop/
27. NASA Science Mission Directorate Artificial Intelligence (SMD AI) Workshop. (2024). Retrieved from https://nasa-impact.github.io/nasa-smd-ai-workshop-2024/
28. Lewis, A.S., Rollinson, C.R., Allyn, A.J., Ashander, J., Brodie, S., Brookson, C.B., Collins, E., Dietze, M.C., Gallinat, A.S., Juvigny-Khenafou, N.: The power of forecasts to advance ecological theory. Methods Ecol. Evol. **14**, 746–756 (2023)
29. Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., Mozaffari, A., Stadtler, S.: Can deep learning beat numerical weather prediction? Philos Trans R Soc A **379**, 20200097 (2021)
30. Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., Palmer, T.: Machine learning emulation of gravity wave drag in numerical weather forecasting. J Adv Model Earth Syst **13**, e2021MS002477 (2021)
31. Grönquist, P., Yao, C., Ben-Nun, T., Dryden, N., Dueben, P., Li, S., Hoefler, T.: Deep learning for post-processing ensemble weather forecasts. Philos Trans R Soc A **379**, 20200092 (2021)
32. Jiang, S., Zheng, Y., Solomatine, D.: Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning. Geophys Res Lett **47**, e2020GL088229 (2020)
33. Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L.: Physics-informed machine learning. Nature Reviews. Physics **3**, 422–440 (2021)
34. An, D., Choi, J.H., Kim, N.H.: Options for prognostics methods: a review of data-driven and physics-based prognostics. In: 54th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, p. 1940 (2013)
35. Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V.: Machine learning for the geosciences: Challenges and opportunities. IEEE Trans. Knowl. Data Eng. **31**, 1544–1554 (2018)
36. Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F.: Scientific machine learning through physics–informed neural networks: Where we are and what's next. J. Sci. Comput. **92**, 88 (2022)
37. Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., et al.: Fourcastnet: a global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint arXiv:2202.11214 (2022)
38. ESA: Using AI to predict Earth's future, https://www.esa.int/Space_Safety/Using_AI_to_predict_Earth_s_future
39. Aristo: SciQ Dataset, https://allenai.org/data/sciq, (2017)
40. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems. 30, (2017)
41. Chen, B., Huang, K., Raghupathi, S., Chandratreya, I., Du, Q., Lipson, H.: Automated discovery of fundamental variables hidden in experimental data. Nature Computational Science. **2**, 433–442 (2022)
42. Badr, H.S., Zaitchik, B.F., Guikema, S.D.: Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel. J. Appl. Meteorol. Climatol. **53**, 614–636 (2014)
43. Rivero-Calle, S., Gnanadesikan, A., Del Castillo, C.E., Balch, W.M., Guikema, S.D.: Multidecadal increase in North Atlantic coccolithophores and the potential role of rising CO2. Science **350**, 1533–1537 (2015)
44. Tripathi, S., Srinivas, V.V., Nanjundiah, R.S.: Downscaling of precipitation for climate change scenarios: a support vector machine approach. J. Hydrol. **330**, 621–640 (2006)
45. Guikema, S.D., Nateghi, R., Quiring, S.M., Staid, A., Reilly, A.C., Gao, M.: Predicting hurricane power outages to support storm response planning. Ieee Access. **2**, 1364–1373 (2014)
46. Guikema, S.: Artificial intelligence for natural hazards risk analysis: Potential, challenges, and research needs. Risk Anal. **40**, 1117–1123 (2020)
47. Stall, S., Cervone, G., Coward, C., Cutcher-Gershenfeld, J., Donaldson, T.J., Chris Erdmann, R., Hanson, B., et al.: Ethical and responsible use of AI/ML in the earth, space, and environmental sciences. Authorea Preprints (2023)

48. Laakasuo, M., Herzon, V., Perander, S., Drosinou, M., Sundvall, J., Palomäki, J., Visala, A.: Socio-cognitive biases in folk AI ethics and risk discourse. AI and Ethics. **1**, 593–610 (2021)
49. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artif. Intell. **267**, 1–38 (2019)
50. Mehrabian, A., Gerasimov, I., Khayat, M., Wei, J., KC, B., Hegde, M.: Natural Language Processing Support for Earth Data Search Engines. In: AGU Fall Meeting Abstracts. pp. IN55A-03 (2021)
51. Lynnes, C., Zhu, M.Q., Blythe, J., Williamson, T.N., Burnett, J., Huffer, E., Armstrong, E.M., Munroe, J.R., Siarto, J., Reese, M.: Usage-Based Discovery of Earth Observations. In: AGU Fall Meeting Abstracts. pp. IN012–02 (2020)
52. Crosby, N., Stoyanova, K., Dayal, R., Gerasimov, I., Mehrabian, A., Wei, J., Pham, L., Khayat, M., Jahoda, E.: Development of a Knowledge Graph for Dataset Discovery and Identification at a NASA Data Center. In: AGU Fall Meeting 2021 (2021)
53. Sun, Z., Cristea, N.C., Yang, K., Alnuaim, A., Bikshapathireddy, L.C.G., John, A., Pflug, J., Li, B., Pan, H., Shyamsunder, N., Reddygari, R., Bhandaru, P.: Making machine learning-based snow water equivalent forecasting research productive and reusable by Geoweaver, Abstract (IN23A-04), presented at AGU Fall Meeting 2022, Chicago, IL, 12-16 December 2022 (2022)
54. Microsoft: Machine Learning Operations, https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/ready/azure-best-practices/ai-machine-learning-mlops
55. Martineau, P.: Toronto tapped artificial intelligence to warn swimmers. The experiment failed. The Information. (2022)
56. Koehl, D.: Finding Needles in a Satellite Imagery Haystack, https://www.earthdata.nasa.gov/learn/articles/spaceml-impact-blog
57. The White House Office of Science and Technology Policy (OSTP). Blueprint for an AI bill of rights: making automated systems work for the American people (2022). Retrieved from https://www.whitehouse.gov/ostp/ai-bill-of-rights/
58. Chollet, F.: Deep learning with Python. Simon and Schuster (2021)
59. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence. **1**, 206–215 (2019)
60. U.S. Government Accountability Office (GAO). GAO-21-519SP highlights, artificial intelligence: an oversight framework for federal agencies and other entities (2021). Retrieved from https://www.gao.gov/assets/gao-21-519sp-highlights.pdf
61. Durbin, C., Quinn, P., Shum, D.: Task 51-cloud-optimized format study. No. GSFC-E-DAA-TN77973 (2020). Retrieved from https://ntrs.nasa.gov/citations/20200001178
62. Maskey M.: NASA earth science data systems/NASA headquarters, interagency implementation and advanced concepts team (IMPACT)/NASA Marshall space flight center. In: Advancing AI for earth science: a data systems perspective. Presented at ESA EO Φ-week (2020). Retrieved from https://ntrs.nasa.gov/api/citations/20205007628/downloads/phi-week.9.28.2020-MM-STI.pdf
63. Jain, S., Mindlin, J., Koren, G., Gulizia, C., Steadman, C., Langendijk, G.S., Osman, M., Abid, M.A., Rao, Y., Rabanal, V.: Are we at risk of losing the current generation of climate researchers to data science? AGU Adv. **3**(4), e2022AV000676 (2022)
64. Salveson, C.: The AI maturity framework. (2020). Retrieved from http://corysalveson.com/publication/eai-ramakrishnan-salveson-2020-ai-maturity-framework/EAI-Ramakrishnan-Salveson-2020-AI-maturity-framework.pdf
65. Lavin, A., Gilligan-Lee, C.M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Baydin, A.G., Sharma, A., Gibson, A.: Technology readiness levels for machine learning systems. Nat. Commun. **13**, 6039 (2022)
66. Mankins, J.C.: Technology readiness levels. White Paper (1995). Retrieved from https://aiaa.kavi.com/apps/group_public/download.php/2212/TRLs_Mankins%20Paper_1995.pdf
67. NASA Earthdata. Data processing levels (2023). Retrieved from https://www.earthdata.nasa.gov/engage/open-data-services-and-software/data-and-information-policy/data-levels
68. Hutson, M.: Artificial intelligence faces reproducibility crisis. Science. **359,** 725–726 (2018)
69. Tseng, G., Kerner, H., Nakalembe, C., Becker-Reshef, I.: Annual and in-season mapping of cropland at field scale with sparse labels. In: Proceedings of the thirty-fourth conference on neural information processing systems workshops (2020)
70. Mozaffari, A., Langguth, M., Gong, B., Ahring, J., Campos, A.R., Nieters, P., Escobar, O.J.C., Wittenbrink, M., Baumann, P., Schultz, M.G.: HPC-oriented canonical workflows for machine learning applications in climate and weather prediction. Data Intelligence. **4**, 271–285 (2022)
71. Kale, A., Nguyen, T., Harris, F.C., Jr., Li, C., Zhang, J., Ma, X.: Provenance documentation to enable explainable and trustworthy AI: A literature review. Data Intelligence. **5**, 139–162 (2023)
72. Binkowski, F.S., Roselle, S.J.: Models-3 community multiscale air quality (CMAQ) model aerosol component 1. Model description. J Geophys Res: Atmos. **108**(D6) (2003)
73. Sun, Z., Di, L., Burgess, A., Tullis, J.A., Magill, A.B.: Geoweaver: Advanced cyberinfrastructure for managing hybrid geoscientific AI workflows. ISPRS Int. J. Geo Inf. **9**, 119 (2020)
74. Sharma, D., Mauzerall, D.: others: Analysis of air pollution data in India between 2015 and 2019. Aerosol and Air Quality Research. **22**, 210204 (2022)
75. CMAQ AI Ozone Forecasting, http://geobrain.csiss.gmu.edu/cmaq_site
76. United States: National oceanic and atmospheric administration. In: NOAA artificial intelligence strategic plan 2021—2025 (2021). https://doi.org/10.25923/fzhw-d487
77. Pedersen, M., Haurum, J.B., Gade, R., Moeslund, T.B.: Detection of marine animals in a new underwater dataset with varying visibility. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 18–26 (2019)
78. Islam, A., Long, C., Basharat, A., Hoogs, A.: DOA-GAN: Dual-order attentive generative adversarial network for image copy-move forgery detection and localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4676–4685 (2020)
79. VIAME: A do-it-yourself AI system for analyzing imagery and video, https://www.viametoolkit.org
80. National Oceanic and Atmospheric Administration, Office for Coastal Management: C-CAP High-Resolution Land Cover and Change, https://coast.noaa.gov/digitalcoast/data/ccaphighres.html, (2016)
81. Chen, Q., Beijbom, O., Chan, S., Bouwmeester, J., Kriegman, D.: A new deep learning engine for CoralNet. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 3693–3702 (2021)
82. CorelNet: Using Machine Learning to Automate Benthic Transect Image Analysis, https://www.livingoceansfoundation.org/science/scientific-collaboration/coralnet/
83. Petrelli, M.: Machine learning in petrology: state-of-the-art and future perspectives. J. Petrol. **65**(5), egae036 (2024)
84. Jorgenson, C., Higgins, O., Petrelli, M., Bégué, F., Caricchi, L.: A machine learning-based approach to clinopyroxene thermobarometry: Model optimization and distribution for use in Earth sciences. J Geophys Res: Solid Earth **127**, e2021JB022904 (2022)
85. Valade, S., Ley, A., Massimetti, F., D'Hondt, O., Laiolo, M., Coppola, D., Loibl, D., Hellwich, O., Walter, T.R.: Towards global volcano monitoring using multisensor sentinel missions

and artificial intelligence: The MOUNTS monitoring system. Remote Sensing. **11**, 1528 (2019)
86. Ren, C.X., Peltier, A., Ferrazzini, V., Rouet-Leduc, B., Johnson, P.A., Brenguier, F.: Machine learning reveals the seismic signature of eruptive behavior at Piton de la Fournaise volcano. Geophys Res Lett **47**, e2019GL085523 (2020)
87. Maeda, Y., Yamanaka, Y., Ito, T., Horikawa, S.: Machine-learning-based detection of volcano seismicity using the spatial pattern of amplitudes. Geophys. J. Int. **225**, 416–444 (2021)
88. Falcin, A., Métaxian, J.-P., Mars, J., Stutzmann, É., Komorowski, J.-C., Moretti, R., Malfante, M., Beauducel, F., Saurel, J.-M., Dessert, C.: A machine-learning approach for automatic classification of volcanic seismicity at La Soufrière Volcano, Guadeloupe. J. Volcanol. Geoth. Res. **411**, 107151 (2021)
89. Ardid, A., Dempsey, D., Caudron, C., Cronin, S.: Seismic precursors to the Whakaari 2019 phreatic eruption are transferable to other eruptions and volcanoes. Nat. Commun. **13**, 2002 (2022)
90. Villarreal, S., Vargas, R.: Representativeness of FLUXNET sites across Latin America. J Geophys Res: Biogeosci **126**, e2020JG006090 (2021)
91. Meyer, H., Pebesma, E.: Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. **12**, 1620–1633 (2021)
92. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D., Gebru, T.: Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency, pp. 220–229 (2019)
93. McRoberts, R.E.: Satellite image-based maps: Scientific inference or pretty pictures? Remote Sens. Environ. **115**, 715–724 (2011)
94. Bates, S., Hastie, T., Tibshirani, R.: Cross-validation: what does it estimate and how well does it do it? J. Am. Stat. Assoc. **119**(546), 1434–1445 (2024)
95. Stehman, S.V., Foody, G.M.: Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. **231**, 111199 (2019)
96. Wadoux, A.M.-C., Heuvelink, G.B., De Bruin, S., Brus, D.J.: Spatial cross-validation is not the right way to evaluate map accuracy. Ecol. Model. **457**, 109692 (2021)
97. Riemann, R., Wilson, B.T., Lister, A., Parks, S.: An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data. Remote Sens. Environ. **114**, 2337–2352 (2010)
98. Lei, J., Wasserman, L.: Distribution-free prediction bands for non-parametric regression. J. R. Stat. Soc. Ser. B Stat Methodol. **76**, 71–96 (2014)
99. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al.: The FAIR guiding principles for scientific data management and stewardship. Sci. Data. **3**(1), 1–9 (2016)
100. Rawal, A., McCoy, J., Rawat, D.B., Sadler, B.M., Amant, R.S.: Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives. IEEE Transactions on Artificial Intelligence. **3**, 852–866 (2021)
101. Peng, G., Lacagnina, C., Downs, R.R., Ganske, A., Ramapriyan, H.K., Ivánová, I., Wyborn L., et al.: Global community guidelines for documenting, sharing, and reusing quality information of individual digital datasets (2022)
102. Ramapriyan, H.K., “Rama”, Leonard, P.J.T.: Data product development guide (DPDG) for data producers version 1 (2020)
103. Dueben, P.D., Schultz, M.G., Chantry, M., Gagne, D.J., Hall, D.M., McGovern, A.: Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook. Artificial Intelligence for the Earth Systems. **1**, e210002 (2022)
104. Loft, R.: Earth system modeling must become more energy efficient. Eos, https://doi.org/10.1029/2020EO147051. (2020)
105. Hodges, B.R., Liu, F.: Integrating AI with physics-based hydrological models and observations for insight into changing climate and anthropogenic impacts. No. AI4ESP-1059. In: Artificial intelligence for earth system predictability (AI4ESP) collaboration (United States) (2021)
106. Kumar, J., Hoffman, F., Rowland, J.: Representing the unrepresented impact of river ice on hydrology, biogeochemistry, vegetation, and geomorphology: a hybrid physics-machine learning approach. No. AI4ESP1073. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States); Los Alamos National Lab. (LANL), Los Alamos, NM (United States) (2021)
107. Zhu, Y., Zhang, R.-H., Moum, J.N., Wang, F., Li, X., Li, D.: Physics-informed deep-learning parameterization of ocean vertical mixing improves climate simulations. National Science Review. 9, nwac044 (2022)
108. Treveil, M., Omont, N., Stenac, C., Lefevre, K., Phan, D., Zentici, J., Lavoillotte, A., Miyazaki, M., Heidmann, L.: Introducing MLOps. O'Reilly Media (2020)
109. Vlaskovits, P.: Henry Ford, innovation, and that “faster horse” quote. Harv. Bus. Rev. **29**, 2011 (2011)
110. Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., others: NetCDF Climate and Forecast (CF) metadata conventions, (2003)
111. Munasinghe, T., Svirsky, A.: Engaging Students in Data Literacy: Lessons Learned from Data Intensive Classrooms. In: Companion Publication of the 13th ACM Web Science Conference 2021. pp. 40–43 (2021)
112. Cristea, N.C., Sun, Z., Arendt, A.A., Henderson, S.T., Denolle, M., Burgess, A.: GeoSMART: Machine Learning Training and Curriculum Development for Earth Science Studies. In: AGU Fall Meeting Abstracts. pp. ED22B-0550 (2022)
113. Huppenkothen, D., Arendt, A., Hogg, D.W., Ram, K., VanderPlas, J.T., Rokem, A.: Hack weeks as a model for data science education and collaboration. Proc. Natl. Acad. Sci. **115**, 8872–8877 (2018)