Report on NSF Sponsored Workshop on AI-Enabled Scientific Revolution

March 8-9, 2023, Alexandria, VA

Authors:

Vipin Kumar (UMN), Aidong Zhang (UVA)

Anuj Karpatne (VT), Parisa Rashidi (UFL), Hongning Wang (UVA)

Milad Abolhasani (NCSU), Chid Apte (IBM Research), Arindam Banerjee (UIUC), Denis Boyda (MIT), Mauro Calabrese (UNC), Nitesh Chawla (Notre Dame), Rama Chellappa (JHU), David Forsyth (UIUC), Ian Foster (UChicago/Argonne), Yolanda Gil (USC), Carla Gomes (Cornell), Jiawei Han (UIUC), Paul Hanson (UWisc), George Karypis (UMN/Amazon), Henry Kautz (Schmidt Futures), Yan Liu (USC), Ioannis Paschalidis (BU), Krishna Rajan (UB), Jordan Read (CUAHSI), Russell Schwartz (CMU), Amarda Shehu (GMU), Alex Smola (Boson), Wei Wang (UCLA).

Section 1. Introduction

Motivation: The past decade has been an inspiring time for artificial intelligence (AI) research. AI systems have transformed the norms and practices across industries and permeated the fabric of human society. Moreover, AI is ushering in a transformative technological age by making remarkable breakthroughs in a number of scientific fields such as protein structure prediction and medical imaging [1-3]. There is an increasing consensus in the wider scientific community that AI is poised to disrupt science by unlocking entirely new approaches, driving new scientific inquiry, and enabling greater scientific leaps with far-reaching societal consequences (e.g., see [26-28]). However, there are substantial barriers preventing science from realizing that potential, and addressing these barriers will require support for advances in AI methods and the adoption of these methods in routine scientific research.

Gaps and Challenges: Current state-of-the-art AI approaches rely primarily on training systems based on big data corpora. For instance, deep learning systems must be shown many examples before they can classify things accurately (in comparison to humans who can learn from only a handful of examples). AI systems can be very brittle; small changes to their operating environment can lead to catastrophic failures, and they can even be exploited through adversarial attacks. Another challenge arises from the fact that real-world scientific data often covers only a limited spectrum of the full range of possible data distributions. Hence, a black-box AI model can perform quite poorly on out-of-sample distributions, even if it happens to show high accuracy on labeled data used for cross-validation [8-10]. Furthermore, in scientific domains, there is discomfort in handing over the analysis to a black box that is hard to interpret despite better predictions. For credible progress, machine learning (ML) models need to transparently integrate scientific domain knowledge or improve existing mechanistic models, which already incorporate the domain knowledge [10-11]. In addition, AI models not grounded in underlying scientific theories are susceptible to producing solutions inconsistent with existing scientific knowledge. In extreme cases, this can create a lack of alignment between the intended behavior and the model's actual performance. A critical challenge is to ensure that the deployed AI system will make correct, interpretable, safe, fair, and reliable decisions.

What is Needed to Address these Gaps: Thus, there is a critical need to catalyze and forge the directions of the next phase of AI systems to address these limitations and to rethink how the sciences engage with AI systems. These systems need to go beyond the fixed training/inference paradigm and be able to provide contextual reasoning and become effective problem-solving partners for us in all situations. Instead of solely relying on big data, a combination of scientific knowledge and common sense driven AI coupled with real-time data will be the key components of the AI decision-making systems of the future. Just as the deep learning springboard opened the door for AI to many new applications in the last ten years, such foundational and use-inspired advances in AI will pave the way for the AI-enabled scientific revolution in the next decade and beyond. Likewise, the practice of science will need to adapt to be able to capitalize on the many ways AI can advance scientific discovery.

Workshop Objectives: A two-day workshop on AI-enabled scientific revolution was held during March 8-9, 2023, at NSF headquarters in Alexandra, Virginia. It brought together leading experts from AI and various scientific and engineering fields (including computational biology, health sciences, neuroscience, chemical and materials science, ecology, climate science, hydrology, limnology, and physics) to identify key challenges and steps that can be taken to enable the next AI revolution in the Sciences. The workshop led the discussion on a new frontier in AI, where novel AI frameworks will drive scientific inquiry, suggest novel experiments, elucidate new theories, and thus revolutionize the traditional discovery process across multiple scientific disciplines. It identified limitations of the current state-of-the-art in AI for existing grand scientific challenge problems and discussed AI advances that are needed to catalyze synergistic research across scientific communities and how these advances can be incorporated into the practice of scientific discovery. The workshop provided an opportunity to have an open conversation with a reflection on the past, present, and future of AI-enabled scientific discoveries. Through the interactions among the participants, the workshop provided an opportunity to align future objectives and new strategies for AI research.

Workshop Structure: Workshop included 28 participants from academia, industry, and philanthropic organizations. In addition, the workshop was attended by around 40 program directors from NSF and other government agencies (including NITRD, NIH, FDA, HHS, DHS). The workshop program included 12 plenary lightning talks held during four sessions and two panels that alternated between AI centric and Science centric discussions. Figure 1 shows the structure of the workshop.

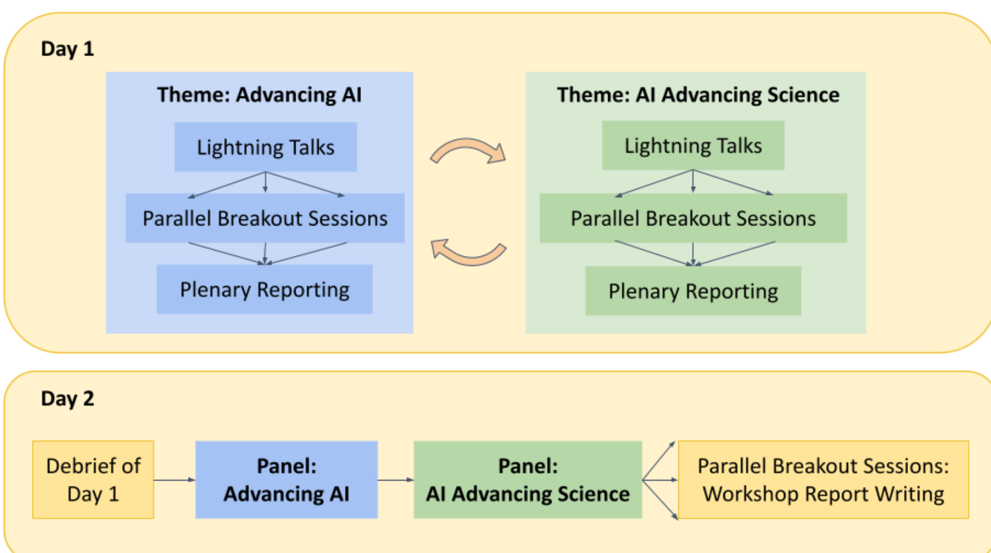

Figure 1. Workshop Structure.

In detail, Day 1 of the workshop program held a series of theme-focused sessions that alternated between AI advances and gaps/needs of the scientific applications. In each session, selected experts were asked to give lightning talks to set the stage, which were followed by discussions in breakout groups to promote discussion in smaller group settings. Each breakout group was assigned a scribe (who kept notes in a Google Doc) and a session lead (who moderated the discussion). Summary of the break-out group discussions were presented to the entire group at the end of each session. Attendees were also able to contribute to the notes for each session via Google docs. The Day 2 AM session started with a summary of the key issues identified during Day 1 AM and PM sessions, and continued with two panels that were devoted to narrowing down the list of recommendations. Day 2 PM session was focused on preparation of the report draft, and ended with concluding remarks from NSF leadership. Additional details are available at the workshop website (https://sites.google.com/umn.edu/nsfaiworkshop2023/).

Outline of the report: As summarized in Figure 2, this report attempts to capture the wide ranging discussions at the workshop on the barriers faced by existing AI, potential advances in AI impacting multiple domains, as well as some concrete recommendations that the workshop participants considered critical for advancing the role of AI in scientific discovery. Specifically, Section 2 examines the current barriers that exist in the application of AI to science. Addressing these barriers will require advances in AI and changes in how scientific research engages with AI technologies. Section 3 discusses the potential AI advances that will impact multiple scientific domains while addressing barriers identified in Section 2. Section 4 offers recommendations to incentivize the development of next-generation AI and its adoption in scientific practice that will dramatically accelerate scientific discovery across a range of domains.

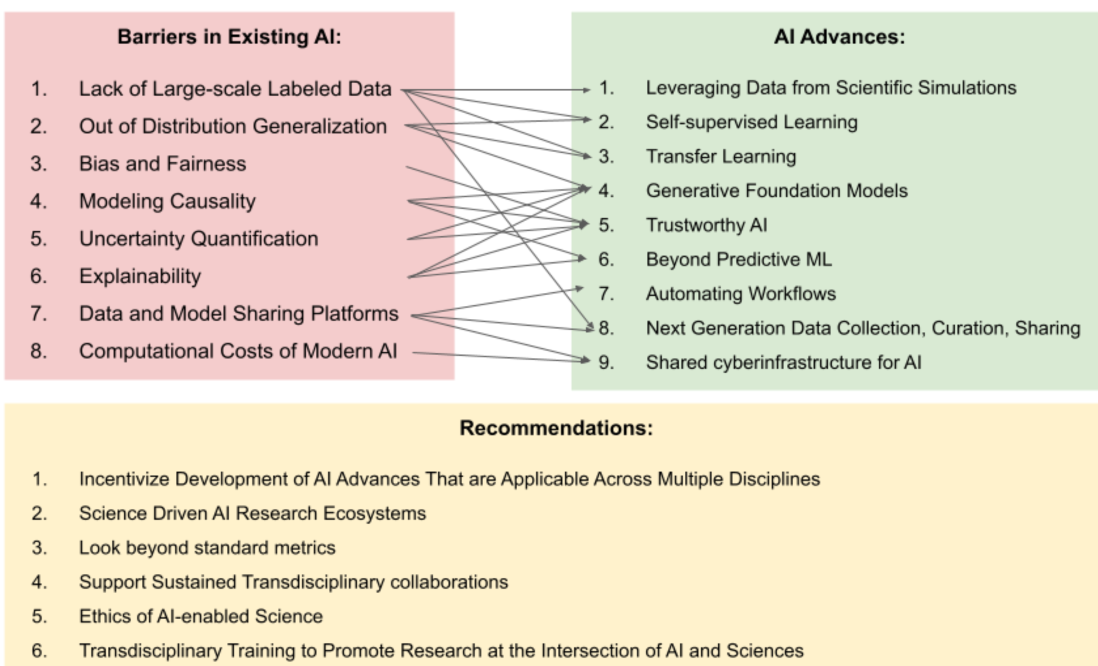

Figure 2. A summary of barriers in existing AI, potential advances in AI impacting multiple domains, and recommendations to support AI for science.

Section 2. Barriers Faced by Existing AI in Science

There are several barriers limiting the widespread adoption of existing AI methods in science, as identified in the following.

2.1. Lack of Large-scale Labeled Data

The scale of labeled samples in scientific domains is often small, covering only a limited spectrum of the full range of possible data distributions. This makes it difficult for state-of-the-art AI/ML methods to achieve good generalization performance, especially on out-of-sample distributions [29] encountered during testing. This is in contrast to the amount of labeled data available in many mainstream applications of computer vision and language modeling, where AI/ML methods have made remarkable progress. As an example, deep learning models for image recognition tasks can require millions of (labeled) images [30], which can be costly and time-consuming to collect and annotate in scientific applications. As a consequence, data-driven AI solutions are known to be prone to overfitting, i.e., capturing spurious correlations in the training data that can lead to poor performance on new data and hinder the generalizability and reproducibility of the AI models. Overfitting is particularly problematic in scientific data, where the number of features and variables can be large but the signal-to-noise ratio can be low. Besides, scientific data may contain outliers, missing values, and measurement errors, which can further increase the chance of overfitting. Scientific data are also available in multiple modalities (e.g., graphs, raster images, and sequence data), at varying scales of observation or simulation and with different levels of accuracy and completeness, making it challenging to apply existing AI methodologies that usually work with a single data modality or two modalities (e.g., images + text [31]). Moreover, while we have ample unlabeled data in some disciplines such as climate science, other fields such as chemistry and material science suffer from sparsity of even unlabeled samples, making it difficult to train traditional unsupervised/semi-supervised learning methods.

2.2. Out-of-distribution Generalization Guarantees

We are interested in building AI/ML models that are robust to distributional shifts of test data compared to training data [29], and even adversarial attacks to the model through the deliberate and optimized perturbation of the inputs [32,33]. We also want to ensure the algorithm's behavior falls within acceptable bounds, minimizing risks and unwanted consequences. However, most AI algorithms are known victims of out-of-distribution samples and adversarial attacks, which cost their robustness and safety guarantees. Specifically, out-of-distribution samples refer to data samples that are different from the training data distribution [5,29]. Most AI models are built on the premise of identical and independent data distribution; but when AI algorithms encounter out-of-distribution samples, they may not be able to make consistent or accurate predictions, leading to poor performance and potential safety concerns. Adversarial attacks are intentional perturbations made to input data (during training or testing) with the aim of deceiving the algorithm and causing it to misclassify the data [32,33]. For instance, adding small amounts of noise to an image can cause an image classifier to identify it as something else entirely. Such attacks pose a significant risk to the safety and reliability of AI in science, particularly in areas where safety is critical, such as medical diagnosis or autonomous vehicles.

Although numerous efforts have been devoted in the AI community on detecting and rejecting out-of-distribution inputs [29,34], incorporating robustness constraints during training [32,33,35], and safety measures [36,37], they cannot cover all potential edge cases or unknown scenarios, especially when algorithms are used in complex systems in science domains. It is also difficult to predict the long-term behavior and impact of an AI system as it learns and adapts to new data gradually. The lack of a comprehensive evaluation framework for assessing the robustness and safety of AI algorithms further poses a significant challenge in this regard. In particular, the lack (or even perception of lack) of trustworthiness of AI models is a significant barrier to their wider use in science. Furthermore, in scientific domains, the behavior of a system is often governed by fundamental laws and principles that may be difficult to capture purely through data. In these cases, it becomes necessary to incorporate domain-specific knowledge and understanding into AI models in order to accurately capture the underlying physics or system principles.

2.3. Bias and Fairness

Bias in AI algorithms refers to the systematic errors or inaccuracies that arise in their output due to various factors, such as the training data, algorithm design, or societal and cultural factors. A model may exhibit bias either because of lack of sufficient training data from specific groups (e.g., underrepresented minorities) or algorithmic biases that arise because the training data may miss hidden variables that could control for these biases [38,39]. We need to account for both types of bias in AI/ML solutions, leading to novel advances in "Fairness in AI." The issue of bias in AI algorithms is particularly important in science domains because scientific research often aims to be objective and unbiased. For instance, if a medical dataset used to train an AI model disproportionately includes data from certain populations or geographic locations, the model may not generalize well to other populations or locations and may produce biased results [40]. Furthermore, there may be unintended biases in data that are not solely dependent on the sampling procedure, e.g., genetic and environmental influences of a disease that favor certain patient groups. As a result, the impact of bias in AI algorithms can be significant, beyond the accuracy of their results, and has the potential to amplify existing biases and discrimination.

2.4. Modeling Causality

A major challenge to the application of AI in science is the limitations of existing AI solutions in discerning causality [4,14]. Most AI algorithms, such as deep learning models, are based on statistical and optimization methods that are designed to identify correlations and associations between input and output (e.g., often formed as an empirical loss minimization problem). These algorithms have shown remarkable success in various applications, such as image recognition, natural language processing, and recommendation systems. However, they are limited by their inability to understand causality.

Understanding causality is crucial in science, as it enables scientists to understand predictions, discover new knowledge, and develop interventions based on the underlying mechanisms. Scientific discovery cannot just rely on statistical associations. For example, in medical research, understanding the causal relationships between risk factors, biomarkers, and diseases can help develop effective treatments and prevention strategies. In environmental science, understanding the causal relationships between pollutants, climate change, and biodiversity loss can help inform policy decisions and conservation efforts. In chemistry, it is important to not only predict chemical reaction outcomes, but also infer synthesis pathways that can more efficiently lead to discovery of new molecules.

A major challenge in causal modeling comes from the presence of confounding variables, where the observed associations between variables may be spurious and not reflect true causal relationships [14]. Confounding variables can arise due to various problem-specific factors, such as unmeasured or hidden variables, selection bias, and reverse causation. State of the art AI methods are limited in automatically identifying the confounders, and domain knowledge is specifically needed to alleviate the challenge. Another challenge is the limited sample size and data availability, which can lead to low statistical power and difficulties in estimating causal effects. In addition, many scientific domains involve complex and dynamic systems, where the causal relationships may be high dimensional, nonlinear, time-varying, and context-dependent. Further, the lack of experimental control and the presence of feedback loops and feedback mechanisms can complicate the causal inference process and make it difficult to distinguish between cause and effect. This is particularly relevant in science fields such as ecology, where the effects of interventions can have indirect and delayed consequences.

2.5. Uncertainty Quantification

A key requirement for AI to have an impact on science is to be able to characterize uncertainty [6], since it is known that AI models may produce results that are over-confident but wrong. Uncertainty characterization and quantification is necessary in scientific settings as the processes and phenomena of interest have aleatoric (statistical) uncertainty, which cannot be reduced by better modeling (which can reduce epistemic uncertainty). For example, take the problem of modeling Atlantic hurricanes. While we cannot realistically say exactly where the eye of the hurricane will be in 6 hours, we need to forecast a suitably calibrated (spatial) probability distribution for policy makers to make decisions. Also, while uncertainty is traditionally categorized as epistemic/aleatoric based on whether the source of uncertainty is reducible or irreducible, there are more ways to describe and categorize sources of uncertainty in scientific problems, e.g., coming from data (stochastic variations or measurement noise), coming from ML model (unknown parameters), or coming from process-based model simulation [10]. More generally, since science often deals with coupled processes and complex interactions, we need models with the ability to not only characterize but propagate such uncertainty. Existing and often popular work on using ensembles in AI to characterize uncertainty have resulted in over-confident models, i.e., underestimation of uncertainty, which can have major undesirable consequences in science. In real-world problems involving large-scale ML models (e.g., foundation models), ensemble methods can still fail as multiple models may be making the same mistakes, even though their variance may be low.

2.6. Explainability

Explainability refers to the ability to understand and interpret how an AI algorithm arrived at its output or decision and thus provide to human users some explanation and intuition about its output [15]. Achieving explainability in AI algorithms is critical for ensuring transparency and accountability in scientific research and for making AI models more trustworthy and widely adopted [8,9]. This is important, particularly in decision-making applications not only to “check the model,” but also to enable a human decision-maker to take into account factors and context that may have not been captured in the data used to train the model. Explainability of results can be a particular concern in scientific contexts because likely users frequently lack computational training themselves and so cannot critically evaluate the AI models. While there are no guarantees that all scientific phenomena are indeed based on first principles, having an intelligible explanation to what the AI methods are doing would be essential in scientific domains.

One of the primary challenges in achieving explainability in AI algorithms is the complexity of modern AI models [41]. Deep learning models, for example, can have billions of parameters and hundreds of layers, making it difficult to understand how their outputs are obtained. This lack of interpretability can be problematic, especially when decisions made by AI models impact human lives or sensitive data. Another challenge in developing explainable AI is the lack of a well-defined and widely accepted evaluation metric for explainability. Unlike other metrics such as accuracy or precision, which have well-defined methods for prediction tasks, there is no universally accepted measure of explainability. This makes it difficult to compare different approaches to explainability or to determine whether a given approach for explanation is effective in a given context. One of the key reasons for the lack of explanation metrics is that the evaluation of explainability is often subjective and context-dependent. What might be considered an acceptable level of explainability in one application or domain may not be sufficient in another. This makes it difficult to define a single set of criteria that can be used to evaluate the explainability of AI algorithms across all applications and domains.

2.7. Data Sharing Mechanisms and Platforms

A major barrier to incorporating AI more broadly into science is data sharing. While some subfields of science have long embraced the need for data sharing, others lag behind. This can limit the availability of data for training AI algorithms and hinder the development of new AI applications in science. The lack of data sharing mechanisms and tools/platforms can be especially challenging for smaller research teams and organizations that may not have the resources to develop their data-sharing infrastructure. This can lead to a disadvantage for smaller teams in fields where data sharing is becoming increasingly important, such as genomics and drug discovery. While some existing data-sharing platforms, such as the Gene Expression Omnibus (GEO) [16] for genomics data or the Protein Data Bank (PDB) [17] for protein structures, have been successful in promoting data sharing, these platforms are often discipline-specific and may not be suitable to be extended for sharing data across multiple scientific disciplines.

Even where data is made available, lack of broad standards for how such data should be distributed impedes the development of general-purpose and cross-disciplinary AI tools for science, particularly when dealing with large and complex datasets, which are common in scientific applications. Scientific data often comes in various formats, structures, and levels of complexity, making it difficult to integrate and compare data from different sources in an efficient manner [26]. Moreover, without properly defined metadata and standards, it can be challenging to understand and consume data, preventing the development of new AI solutions. For example, different scientific domains often use different terminologies, measurement units, and data structures and formats, making it challenging to integrate data from different sources. However, creating and maintaining metadata can be time-consuming and challenging, requiring significant resources and expertise. Scientific applications can also introduce some compelling reasons to not share data freely, e.g., the need to protect privacy of human research subjects.

Another significant challenge related to data sharing in AI for science lies on its sustainability aspect [42]. While many data sharing efforts and practices exist, they often require significant resources to maintain (including human efforts and sophisticated software systems), update, and ensure the quality of the shared data. Moreover, funding for data sharing initiatives is often limited in general, despite the great effort and investment made by several government funding agencies including NSF, making it challenging to sustain data sharing activities over time. For example, in many cases, data sharing platforms were developed as part of research projects and may not receive continued funding or maintenance after the project has ended. This can lead to a situation where data sharing activities cease or become less effective over time. Furthermore, the sustainability of data sharing can be challenging due to the need for continuous data management practices, such as data annotation, metadata creation, and data quality control. These activities require significant resources and expertise, which may not be available to all research teams.

Finally, much data that is important for research in medicine and biology are subject to human subjects regulations. It can only be shared when appropriate legal agreements regarding the use and protection of data that could be traced back to individuals are made. Data-sharing platforms for such data must not only support storage and distribution of data, but also support the creation and maintenance of data-use agreements between data creators and data users. In some cases, data sets restricted by human subjects rules can be replaced by surrogate data that captures its significant statistical patterns but is fully deidentified. Data sharing platforms need to include tools for creating surrogate data when possible.

2.8 Computational Costs of Modern AI

The cost of modern AI model training and execution can be prohibitively high [19], making it a significant barrier to the widespread adoption of AI in science. In particular, some of the high profile successes of AI algorithms in computer vision and natural language processing would not have been possible without significant computational resources, such as high-performance GPUs and specialized hardware, which can be expensive to acquire and maintain by scientific communities [43]. Simply replicating this process in all science domains is clearly not scalable.

The cost of AI model training and execution can be especially challenging for smaller research teams and organizations, as they may not have access to the necessary data and computational resources to develop and fine-tune models. For example, in the field of drug discovery, AI is being increasingly used to identify potential candidates for new therapies [18]. However, the training of large-scale deep learning models to predict drug-protein interactions requires significant computational resources and vast amounts of data, which can be costly for smaller research teams. In addition to the high cost of data and computational resources, there is also a significant energy cost associated with training and deploying AI models. Deep learning models, for instance, require a large number of computations to be performed in parallel, which can lead to high power consumption. For example, the energy cost of training a single deep learning model can be equivalent to the energy consumption and carbon emission of multiple households over several months [19]. Advances that reduce computational cost of training AI models will make it easier for domain experts across the sciences to apply AI to bespoke problems.

Section 3. Potential AI Advances Impacting Multiple Scientific Domains

The workshop provided the opportunity to discuss AI advances that are needed to address the barriers mentioned in Section 2 and have impacts across multiple scientific domains.

A cross-cutting theme in these advances is to incorporate scientific knowledge and structure into AI/ML models. Scientific knowledge and structure may correspond to conservation laws, simulations, ontologies, expert-derived rules, etc. [12, 26]. Using such knowledge can help to constrain the search space of AI/ML solutions to scientifically consistent solutions leading to better out-of-distribution generalizability even in the paucity of labeled data [11, 20, 21]. AI/ML models grounded in scientific knowledge are also more likely to be explainable to domain scientists, produce robust results with better quantification of uncertainties, and lead to the generation of novel scientific hypotheses explaining the causality of relationships inferred from data useful for scientific advancements [10]. This is an iterative process, as we can use AI/ML models to populate scientific knowledge bases and also to leverage knowledge bases to improve ML performance. It was also considered important to develop evaluation metrics that go beyond accuracy and consider factors such as robustness and interpretability, which requires deep understanding of the domain specific properties and reasoning in science.

In the following, we describe some of the prominent AI advances that were discussed at the workshop and the open questions we need to explore in each advance to address the barriers faced by existing AI methods in scientific applications.

3.1. Leveraging Data from Scientific Simulations

A central advance that is needed in AI for widespread adoption in science is the ability to learn from a mix of observational data and simulation data generated by scientific models operating on first principles knowledge. AI models have to consider the fact that the simulation data is possibly based on incomplete scientific knowledge and is not of the same scientific quality as the observational data. Further, scientific knowledge has to be used to enable suitable inductive biases to aid the learning [20]. As an example, consider the problem of forecasting the trajectory of Atlantic hurricanes where state-of-the-art scientific, statistical, and machine learning models have all struggled to provide persistent high quality forecasts. The observational data is limited, simulations are available but arguably still not an accurate scientific or statistical representation of reality, and there is considerable scientific understanding of atmospheric and oceanic processes and their couplings. The central tenet of AI for science holds the promise to be able to use such a mix of small high quality observational data, reasonably large but not as high quality simulation data, and core scientific knowledge to help develop accurate models for challenging problems.

3.2. Self-supervised Learning

To address the barrier of limited large-scale labeled data in scientific domains, there is a growing interest in exploring alternative paradigms in AI such as self-supervised, semi-supervised, distantly-supervised, and weakly-supervised learning. These paradigms rely on other forms of supervision for training ML models that are less reliant on labels. Novel advances are required in these paradigms that take advantage of the unique properties of scientific data, e.g., knowledge of scientific principles driving real-world phenomena. A promising example is the use of scientific knowledge-guided loss functions in the training of deep learning models for a range of applications including lake modeling, fluid dynamics, quantum mechanics, and solving partial differential equations [10, 20]. Some open questions that we need to explore along this AI advance include: how can we use scientific knowledge to inform AI models when we also have uncertainty in domain knowledge? How can models be improved by knowledge-guided reasoning (e.g., by incorporating formal methods/verification into modeling, or by adopting physics-inspired AI model architectures and algorithms)? How can we explore the hard problem of incorporating human knowledge/intuition in the AI process, e.g., problems where answers are constrained by moral issues?

3.3. Transfer Learning

There is a lot of progress that has been made in creating large-scale foundation models for language and vision problems where knowledge from multiple sources and tasks are pooled together to learn generalizable features or representations that can be transferred or fine-tuned to a new task with limited training cycles. Novel AI advances are required to build upon such efforts to create foundation models for a group of related scientific problems or tasks. In many scientific problems, models built for one scenario need to be transferred to other scenarios with varying system characteristics, data distributions, or system sizes. For example, in hydrology and ecology, we are interested in training surrogate models of water quality processes that can be transferred from highly observed systems to less observed or completely unobserved systems [25]. As another example, in condensed matter theory, we are interested in building AI solvers for the wave equation of a system of n particles which when trained on n=10 systems, can generalize even for n>10 system sizes [22]. Since scientific laws can generalize to a range of input/output distributions and system sizes, we would expect AI models in science to exhibit similar types of generalizability.

3.4. Generative Foundation Models

A major advance that is needed in the current state of generative foundation models (e.g., large language models or LLMs) to make an impact in science is to integrate reasoning capabilities in their solution structure, e.g., using recent developments in neuro-symbolic AI [23]. This can enable us to use scientific knowledge available as ontologies (e.g., in biology) to reason about the results of foundation models, leading to better consistency and scientific trustworthiness of results. By building upon recent developments in neuro-symbolic AI, we can constrain structures in neural models with neural reasoning, bridging different perspectives and models in the field of AI. Since neural models are homogenous and tend to make similar mistakes, e.g., when trained on the same data or similar architectures, it is important to promote diversity in training data and model architectures to reduce the risk of common mistakes. We also need to go beyond information extraction and summarization capabilities of foundation models such as chatGPT [24] to create generative models for specific scientific domains that can accelerate data synthesis and enable scientific experimentation. Fact-checking or verification of generative models is another important aspect that needs attention from the AI community. For example, LLMs linked to information sources might also lead to the situation where users deliberately change the sources with downstream impact on the models (injecting misinformation into the models). Instead, we need novel research to link LLMs to correct information sources, which is currently lacking. Some open questions that we need to explore in this AI advance include: can we create generative models that can model complex dynamical systems in science and be useful for applications such as answering what-if scenarios in climate science, generating new materials and drugs that optimize desired properties? How can we model dynamically evolving non-stationary higher order interactions in scientific systems using novel AI advances? Can we reduce the size of large foundation models without affecting their competency for widespread deployment in scientific problems?

3.5. Trustworthy AI

Along with serving the goal of better predictive performance, novel AI advances are needed to solve the end-goal of discovering new scientific knowledge from data. While there is a lot of focus on developing explainable AI approaches in mainstream applications for ensuring fairness and trustworthiness of results, scientific problems may require new forms of explainability. For example, the discovered patterns/insights/rules should be consistent with existing bodies of scientific knowledge and should be explainable to domain scientists, to seed scientific hypotheses that can eventually be scientifically validated and lead to advancement of scientific knowledge. We need to develop capabilities to augment predictive models with appropriate scientific meta-justification, and even go beyond current notions of explainability in mainstream AI/ML applications, since the nature of understanding in scientific discovery is fundamentally different from current standards of explainability tools that attribute input regions associated with the generated predictions. We also need to go beyond modeling associations to understanding the causality of the system, e.g., using counterfactuals. Conceptual frameworks to improve explainability of AI solutions will help us figure out what works and what doesn't work and predict functional outcomes. By going beyond statistical models to causality-based models, we can even go backwards, i.e., perform inverse modeling, where we try to determine some underlying/hidden aspects of a system by just observing its behavior or response to some drivers. AI researchers need to revisit the concept of explainable AI and possibly learn from and interact with scientists in specific science areas to understand their implication, measures and methods for control.

As AI systems become increasingly advanced, it's important for researchers to look beyond simple prediction tasks and explore more complex capabilities, including inverse design, prescriptive modeling, and causal inference. These capabilities have the potential to revolutionize fields like materials science, drug discovery, and climate modeling, allowing researchers to quickly identify optimal designs and solutions with minimal trial and error. To achieve this, future AI systems will need to be "open box", allowing researchers to understand how the system arrived at its recommendations and enabling them to refine and improve the system over time. By embracing these advanced capabilities and ensuring transparency and interpretability, we can unlock the full potential of AI and usher in a new era of discovery and innovation.

Another advance that is needed in the field of AI for science is to create a new generation of AI models that characterize uncertainty accurately in a rigorous sense. We need suitable uncertainty propagation and sampling techniques which support inference in coupled systems. Such developments will possibly be able to leverage advances in generative AI and related themes, especially ideas based on Langevin dynamics, stochastic differential equations, Hamiltonian Monte Carlo, among others. Some of the open questions that we need to explore in this advance include: can we develop theory to understand the limitations of AI models and predict beforehand when AI models will fail and by how much given the characteristics of a new data? Can we make the model self-aware, so that it knows what it does not know?

3.6. Beyond Predictive Machine Learning

While most of the high profile successes in AI have come from predictive machine learning, there is an opportunity to explore a new forms of AI that go beyond statistical predictive modeling, including symbolic (e.g., reasoning and planning), prescriptive modeling (e.g., making recommendations to optimize future outcomes), and numerical (e.g., physics-guided modeling). This may enhance AI capabilities in generating credible hypotheses at scale and enhancing its decision-support functionality. AI models of today typically make one-shot decisions and there is less research on models that make decisions dynamically, over time, when errors get compounded and may lead to catastrophic failures. Extending AI successes to these dynamic settings is an important priority. This might require developing new models to enable scientific discovery, e.g., Natural Language Processing (NLP) tools for identifying knowledge gaps and discovering novel research questions in a reliable and explainable manner. It may also require leveraging methods from dynamical systems and random processes. It would also be essential to develop companion tools for verifying the correctness of model outputs. These tools can assist in various tasks, including AI-assisted construction of knowledge bases from data, cross-domain collaboration for data collection and curation, integrating data from different types and sources, optimizing scientific tools and workflows, AI-assisted evaluation/benchmarking, understanding scientific phenomena, and inverse design.

3.7. Intelligent Automation of Research Workflows

AI can be used as a research assistant to scale routine tasks. AI-assistance in science may also mean curating or discovering datasets and data engineering/plumbing. Scientific work can be scaled beyond the work of human researchers by delegating some of the scientific tasks to AI assistants. AI can be used for several tasks in the scientific discovery process including optimization, decision making, planning, robotics, and search. We need AI methods to support all of these tasks in the scientific process, and not just focus on improving predictive performance. By accelerating workflows using AI, we can potentially reduce labor and cost in science and save efforts in repetitive tasks. However, thinking out of the box of existing knowledge is also important to generate new hypotheses, instead of converging on existing knowledge. Some of the open questions that we need to explore in this advance include: just like autoML can be used to search in hyper-parameter spaces of AI models, can we build AI models that can learn AI tools on their own (e.g., as ToolFormers or Augmented LMs)? How can we perform automated construction of scientific knowledge bases directly from data? How can we define a utility function for a “real-world” problem using AI (e.g., by learning utility functions by querying experts)?

3.8. Next Generation Data Collection, Curation, and Sharing

Collecting, curating, standardizing, and sharing data can lead to large gains in the impact of AI for science compared to other efforts. We need to think of best practices in information collection, and prioritize where we want to get data from for training/testing AI models. A unique opportunity in science is to explore cross-domain collaboration of data collection and curation efforts. For example, cooperative collection of field data by different observing groups can lead to large-scale cost savings in terms of sensor placements or deployment of humans to validate measurements. We also need to optimize tools and workflows for data curation and collection, integrating data from different types and sources, and developing automated evaluation and benchmarking tools and platforms, e.g., self-driving labs. A challenge to be addressed is how do we handle multi-X data (X=modal, resolution, scale, source) with sparse distributions by pooling data from different sources? Can AI help integrate / merge different data types? Easy examples include data from different resolutions for geospatial analysis. Difficult examples include collecting data from different experimental protocols that may be poorly described. Some open questions that we need to explore include: can AI help in data standardization (e.g.,. the data types and data generating processes)? In scientific problems where data can be quite messy, expensive, and of varying types (e.g., from observations and simulations), is there a need to focus on data-centric AI approaches? Are there intelligent ways of selecting samples from observed data or requesting simulations with input parameter configurations that can efficiently learn generalizable AI models in science?

3.9. Shared Cyberinfrastructure for AI

We need advances in cyberinfrastructure to support sharing of data and AI models, as well as in the availability of computing resources to train and evaluate AI models. It is important for both science and AI fields to generalize from successful practices of data sharing in some domains, such as computational biology, to others where such practices are still missing. In addition, regulations and policies that facilitate data sharing, particularly for privacy-critical data (e.g., HIPAA), need to be revisited. There is a need to “strongly encourage” data sharing for de-identified data and reduce the regulatory requirements for such data sharing with the understanding that there is no perfect privacy. Further technical advances in privacy-preserving methods and federated learning are also needed to overcome the barriers in free sharing of data and AI models. A useful analogy may be what federal agencies did with the publishing industry, requiring open access. Is it possible to replicate this model for better data sharing in general science fields? Furthermore, the existing platforms may not be optimized for AI applications, which require large amounts of data, especially labeled data, and may have specific requirements for data quality and metadata. Further efforts are needed to develop more comprehensive and accessible data-sharing mechanisms and tools/platforms that are optimized for AI applications and can support collaboration across multiple scientific disciplines.

Advancing science requires a shared commitment to reciprocity --- those who generate data and depend on data analysis tools to use those data in turn make their data and derivations available for development of the next generation of tools and knowledge. We also need to improve software usability/tooling in AI for science. In other words, we want AI software to be such that they can be used by the entire community in a consistent and reproducible way.

Section 4. Recommendations

This section contains recommendations to incentivize the development of next-generation AI and its adoption in scientific practice that will dramatically accelerate scientific discovery across a range of domains. There was an agreement amongst the workshop participants that challenges unique to scientific problems offer an opportunity to advance AI. In addition, AI advances made in the context of specific scientific domains have the potential to be relevant to many other domains. Section 3 lists a number of areas where AI advances need to be made to address the challenges faced by state-of-the-art AI in scientific applications listed in Section 2. Many of these challenges are already being explored in the context of specific scientific domains (e.g., as part of the AI Institutes and other AI+X programs such as CDS&E). In addition, several challenges identified in Section 2 (e.g., ethics, fairness, trustworthiness) are highly relevant outside of scientific applications.

4.1. Incentivize Development of AI Advances That are Applicable Across Multiple Scientific Disciplines

As AI methods are applied in specific scientific disciplines, new extensions are made in response to real data and real challenges inherent to applications. However, the cross-domain applications of such domain-specific extensions are often not explored. For example, a bioinformatics researcher who has extended a Deep Learning method, may not be interested (i.e., has little incentive and/or background) in investigating whether it can be applied to materials science. While AI+X programs (where X is a scientific discipline) are critical for making domain-inspired advances in AI, such siloed efforts may create solutions that may seem very specifically suited only for the application at hand.

New federal programs that focus on generalization and cross-disciplinary AI applications will be unique and different from the AI institutes. Such programs should aim to fund projects that are not just scaled-up versions of existing AI+X efforts but represent a "*new way of thinking*." These programs should emphasize AI advances that are applicable across multiple scientific disciplines. In particular, they should encourage the transfer of AI advances in one scientific domain to another, or to generalize across many scientific domains. This will be challenging, as even closely related scientific disciplines such as hydrology and limnology, or material and drug design, can have very different cultures and terminologies. It is already challenging for AI researchers to keep pace with the rapidly expanding new vocabulary and specific data processing approaches, tools, baselines, and more in even one scientific discipline. Nevertheless, such programs will amplify what is being accomplished under existing AI+X programs, by taking the lessons learned in one discipline to many other disciplines.

Such programs will pursue methodological advances (as identified in Section 3) that show applicability to more than one scientific domain. Here we list a few illustrative examples. (1) Large language models (LLM) that can be used for scientific hypothesis generation, as LLMs have the ability to analyze vast amounts of scientific literature, text and image data, and other sources to identify patterns and potential relationships between different scientific concepts. This can help scientists in different domains with new insights, connections, and predictions that can help guide further research and experimentation. (2) Explainable AI approaches that can be applied to many scientific domains to ensure fairness and trustworthiness of results from data-driven solutions. (3) Self-supervised learning approaches that can leverage scientific knowledge in the form of conservation laws and/or simulation data from imperfect models that are available in many domains.

4.2. Science-Driven AI Research Ecosystems

We need to develop resources and ecosystems that allow the AI community to explore scientific applications as effortlessly as traditional non-scientific applications. Many of the grand successes of AI and methods development have been done in the context of applications such as computer vision and language processing, and have been enabled by large benchmark datasets and problem formulations developed in these communities. There is a need to develop similar ‘ecosystems’ that are built around scientific applications and can serve as drivers for AI advances. In particular, efforts should be made to identify key data bottlenecks through scientific discipline-specific convenings and conversations between funding agency program directors and scientists. These conversations should identify incentives needed to move the community to meet the need. Funding agencies, both governmental and private, should coordinate their efforts to create new programs and funding mechanisms to provide such incentives.

In addition to creating and curating large data sets, we also need novel formulations and metrics that have utility across multiple scientific disciplines. For example, building the next-generation AI methods for handling data from disparate modalities and scales is considered a major challenge in many scientific applications, where the evolution of natural processes can be observed by disparate instruments (modalities) and the underlying phenomena can be studied at different scales. As another example, inverse problems (where one is required to identify the underlying state of a system by simply observing its output) are some of the hardest problems in many scientific disciplines, and yet they are not suited for the traditional ML-based predictive models that largely focus on forward modeling. Creating novel problem formulations and large scale benchmarks that are relevant for multiple scientific disciplines can usher in a new generation of AI.

4.3. Look Beyond Standard Metrics

Measuring success in many scientific problems is hard to capture via standard performance metrics that are commonly used in the design of predictive ML models. In addition, there is too much focus in the AI community on developing new methods that improve performance relative to state-of-the-art methods on standard benchmarks. We need to incentivize the community to go beyond such incremental improvements on standard benchmarks and focus on measuring impact on real applications and in the real world (e.g., number of patient interventions that could be enabled using a decision support model, quality improvement to human decisions when aided by AI, implementation costs and model performance, as opposed to solely reporting ML performance measures). Each community could develop a new set of domain-specific and tailored metrics building upon generic metrics, potentially resulting in an ontological performance metrics space.

4.4. Support Sustained Transdisciplinary Collaborations

We need to explore a two-way street between AI and domain sciences with sustained transdisciplinary collaborations and support for dedicated experts working at the boundaries between AI and scientific domains. Such collaborations can be strengthened by partnerships amongst academia, industry, and government labs, each bringing its unique strengths and resources. For example, industry often has data that they may not be willing to share publicly but may make it available to a partnership. Government labs such as DOE are able to focus on long term research programs around societal grand challenges. Academic institutions can provide the next generation of researchers and federal agencies could support such collaborations via innovative and collaborative programs.

NSF could incentivize transdisciplinary, AI-centric events that bring scientists together, similar to the current workshop, but on a larger scale. At relatively insignificant costs, such events will allow NSF to bring together funding partners across federal agencies, as well as disparate communities that might otherwise not have many opportunities to communicate and exchange ideas.

Large, regional computing centers that galvanize consortia of universities may be powerful mechanisms to provide industry-scale resources, as well as a sense of urgency and focus that disparate smaller efforts often fail to convey or converge. Such centers may be critical to additionally training the next generation of truly interdisciplinary AI & Science students and researchers.

4.5. Ethics of AI-enabled Science

As AI-enabled science tools will become more commonplace, it will be imperative to also develop an evolving set of guidelines for ethical data collection and synthesis for building AI-enabled science tools. For example, to maintain integrity, data provenance and trustworthy source crediting tools must be developed to track and credit training data. In addition, AI models, particularly those that involve human subjects (e.g., in health applications), need to be vetted to ensure they abide by certain ethical considerations, warranting that the outcomes across different demographic groups, such as different age groups, gender, or races, are fairly distributed. Similar to other AI tools, unintended consequences in science should be recognized and addressed, including dual-use technologies and model manipulation by adversaries to fabricate or misrepresent scientific facts and thus appropriate countermeasures need to be considered.

4.6. Transdisciplinary Training to Promote Research at the Intersection of AI and Sciences

While AI has tremendous potential to improve the practice at every stage of scientific discovery, making use of that potential will require scientists with at least a working knowledge of AI methods to be making decisions at every stage of science from hypothesis generation, through experimental design, execution, and data interpretation. To promote a deeper integration between AI and specific science domains, we need to place more emphasis on interdisciplinary training of students and practitioners. AI for science can serve as a conduit of ideas for various disciplines. Institutions will need to provide domain-specific AI education to trainees in the future to foster better cross-disciplinary communication and collaboration and to prepare them for future workforce needs. For example, future patent/copyright lawyers would need to be familiar with the issues and flaws surrounding various generative models. To this end, training programs for undergraduates and graduate students should be offered to provide structured training in an interdisciplinary manner.

There is a related problem of educating experts in AI sufficiently for them to interact productively with scientific practice. While practicing experimental scientists need to be able to use current AI methods effectively in their labs, improving the state-of-the-art will typically require team science involving experimental domain experts, AI domain experts, and true bridge scientists conversant in the language and problems of both fields. This will require cultivating interdisciplinary educational programs to ensure a critical mass of those bridge experts and reconsidering the education of computational scientists to ensure a pool of AI domain experts with at least a working knowledge of science and the scientific discovery process.

References

[1] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. and Bridgland, A., 2021. Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), pp.583-589.

[2] Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., Tanburn, R., Battaglia, P., Blundell, C., Juhász, A. and Lackenby, M., 2021. Advancing mathematics by guiding human intuition with AI. Nature, 600(7887), pp.70-74.

[3] Aggarwal, R., Sounderajah, V., Martin, G., Ting, D.S., Karthikesalingam, A., King, D., Ashrafian, H. and Darzi, A., 2021. Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. NPJ digital medicine, 4(1), pp.1-23.

[4] Schölkopf, B., Locatello, F., Bauer, S., Ke, N.R., Kalchbrenner, N., Goyal, A. and Bengio, Y., 2021. Toward causal representation learning. Proceedings of the IEEE, 109(5), pp.612-634.

[5] Zhang, C., Bengio, S., Hardt, M., Recht, B. and Vinyals, O., 2021. Understanding deep learning (still) requires rethinking generalization. Communications of the ACM, 64(3), pp.107-115.

[6] Amini, A., Schwarting, W., Soleimany, A. and Rus, D., 2020. Deep evidential regression. Advances in Neural Information Processing Systems, 33, pp.14927-14937.

[7] Zoph, B., Vasudevan, V., Shlens, J. and Le, Q.V., 2018. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8697-8710).

[8] Mengdi Huai, Jinduo Liu, Chenglin Miao, Liuyi Yao, Aidong Zhang, Towards Automating Model Explanations with Certified Robustness Guarantee, Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-2022), Vancouver Convention Centre, Canada, Feb 21-28, 2022.

[9] Sanchit Sinha, Mengdi Huai, Jianhui Sun, and Aidong Zhang, Understanding and Enhancing Robustness of Concept-based Models, the Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-2023), Washington, DC, February 7-14, 2023.

[10] Anuj Karpatne, Ramakrishnan Kannan, Vipin Kumar, Knowledge-Guided Machine Learning: Accelerating Discovery Using Scientific Knowledge and Data, Data Mining and Knowledge Discovery Series of Chapman and Hall/CRC Press, ISBN-9780367693411, 2022.

[11] Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Transactions on knowledge and data engineering, 29(10):2318–2331, 2017.

[12] Willard, J., Jia, X., Xu, S., Steinbach, M. and Kumar, V., Integrating scientific knowledge with machine learning for engineering and environmental systems. ACM Computing Surveys (CSUR), 55(4):1-37, 2022

[13] National Artificial Intelligence Research Resource Task Force, 2023. Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource. Final Report. https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf

[14] Pearl, Judea. "The seven tools of causal inference, with reflections on machine learning." Communications of the ACM 62.3 (2019): 54-60.

[15] Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." Information fusion 58 (2020): 82-115.

[16] Clough, Emily, and Tanya Barrett. "The gene expression omnibus database." Statistical Genomics: Methods and Protocols (2016): 93-110.

[17] Sussman, Joel L., Dawei Lin, Jiansheng Jiang, Nancy O. Manning, Jaime Prilusky, Otto Ritter, and Enrique E. Abola. "Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules." Acta Crystallographica Section D: Biological Crystallography 54, no. 6 (1998): 1078-1084.

[18] Fleming, Nic. "How artificial intelligence is changing drug discovery." Nature 557, no. 7706 (2018): S55-S55.

[19] Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and policy considerations for modern deep learning research." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 09, pp. 13693-13696. 2020.

[20] Karniadakis, George Em, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. "Physics-informed machine learning." Nature Reviews Physics 3, no. 6 (2021): 422-440.

[21] Read, Jordan S., Xiaowei Jia, Jared Willard, Alison P. Appling, Jacob A. Zwart, Samantha K. Oliver, Anuj Karpatne et al. "Process-guided deep learning predictions of lake water temperature." Water Resources Research 55, no. 11 (2019): 9173-9190.

[22] Elhamod, Mohannad, Jie Bu, Christopher Singh, Matthew Redell, Abantika Ghosh, Viktor Podolskiy, Wei-Cheng Lee, and Anuj Karpatne. "CoPhy-PGNN: Learning physics-guided neural networks with competing loss functions for solving eigenvalue problems." ACM Transactions on Intelligent Systems and Technology 13, no. 6 (2022): 1-23.

[23] Sarker, Md Kamruzzaman, Lu Zhou, Aaron Eberhart, and Pascal Hitzler. "Neuro-symbolic artificial intelligence." AI Communications 34, no. 3 (2021): 197-209.

[24] Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.

[25] JD Willard, JS Read, S Topp, GJA Hansen, V Kumar, "Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning (1980–2020)."Limnology and Oceanography Letters 7 (4) (2022), 287-301.

[26] Gil, Yolanda, Suzanne A. Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham et al. "Intelligent systems for geosciences: an essential research agenda." Communications of the ACM 62, no. 1 (2018): 76-84.

[27] Appenzeller, Tim. "The AI revolution in science." Science 357 (2017): 16-17.

[28] Stevens, R., Taylor, V., Nichols, J., Maccabe, A. B., Yelick, K., & Brown, D. (2020). Ai for science: Report on the department of energy (doe) town halls on artificial intelligence (ai) for science (No. ANL-20/17). Argonne National Lab.(ANL), Argonne, IL (United States).

[29] Krueger, D., Caballero, E., Jacobsen, J.H., Zhang, A., Binas, J., Zhang, D., Le Priol, R. and Courville, A., 2021, July. Out-of-distribution generalization via risk extrapolation (rex). In International Conference on Machine Learning (pp. 5815-5826). PMLR.

[30] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[31] Ma, L., Lu, Z., Shang, L. and Li, H., 2015. Multimodal convolutional neural networks for matching image and sentence. In Proceedings of the IEEE international conference on computer vision (pp. 2623-2631).

[32] Kurakin, A., Goodfellow, I.J. and Bengio, S., Adversarial Machine Learning at Scale. In International Conference on Learning Representations.

[33] Chakraborty, A., Alam, M., Dey, V., Chattopadhyay, A. and Mukhopadhyay, D., 2018. Adversarial attacks and defenses: A survey. arXiv preprint arXiv:1810.00069.

[34] DeVries, T. and Taylor, G.W., 2018. Learning confidence for out-of-distribution detection in neural networks. arXiv preprint arXiv:1802.04865.

[35] Fawzi, A., Fawzi, O. and Frossard, P., 2018. Analysis of classifiers' robustness to adversarial perturbations. Machine learning, 107(3), pp.481-508.

[36] Turchetta, M., Berkenkamp, F. and Krause, A., 2019. Safe exploration for interactive machine learning. Advances in Neural Information Processing Systems, 32.

[37] Garcıa, J. and Fernández, F., 2015. A comprehensive survey on safe reinforcement learning. Journal of Machine Learning Research, 16(1), pp.1437-1480.

[38] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A., 2021. A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), pp.1-35.

[39] Yapo, A. and Weiss, J., 2018. Ethical implications of bias in machine learning.

[40] Gianfrancesco, M.A., Tamang, S., Yazdany, J. and Schmajuk, G., 2018. Potential biases in machine learning algorithms using electronic health record data. JAMA internal medicine, 178(11), pp.1544-1547.

[41] Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S., 2020. Explainable ai: A review of machine learning interpretability methods. Entropy, 23(1), p.18.

[42] Kaye, J. and Hawkins, N., 2014. Data sharing policy design for consortia: challenges for sustainability. Genome medicine, 6(1), pp.1-8.

[43] Strubell, E., Ganesh, A. and McCallum, A., 2019, July. Energy and Policy Considerations for Deep Learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645-3650).