ARL-SR-0510 • SEP 2025

Workshop on Cross-Disciplinary Challenges and Opportunities in AI Applications to Science and Engineering (C2OA2SE)

by Sarah Kitchen and Steven Thurman

NOTICES

Disclaimers

ARL-SR-0510 ● SEP 2025

Workshop on Cross-Disciplinary Challenges and Opportunities in AI Applications to Science and Engineering (C2OA2SE)

Sarah Kitchen
Michigan Tech Research Institute

Steven Thurman
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	END DATE
September 2025	Special Report	6 Nov 2024	8 Nov 2024

4. TITLE AND SUBTITLE
Workshop on Cross-Disciplinary Challenges and Opportunities in AI Applications to Science and Engineering (C2OA2SE)

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
W911NF2410227		
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Sarah Kitchen and Steven Thurman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-FE Aberdeen Proving Ground, MD 21005	ARL-SR-0510

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCID: Steven Thurman, 0000-0003-3962-8447

14. ABSTRACT
This report is a summary of the work presented at the Workshop on Cross-Disciplinary Challenges and Opportunities in AI Applications to Science and Engineering (C2OA2SE), which was held to address the integration of AI into diverse scientific and engineering fields. Driven by recent advances in AI algorithms, data availability, and computational resources, the workshop convened 13 invited speakers and 21 additional participants to examine state-of-the-art AI approaches and identify common challenges across disciplines. The workshop objectives centered on fostering cross-disciplinary dialogue, sharing complex problems, and characterizing approaches for rigorous investigation of complex systems leveraging AI. Activities included technical talks, focused breakout sessions, and large group discussions designed to facilitate the exchange of knowledge and best practices. Organized by the Michigan Tech Research Institute and ARL, C2OA2SE highlighted the need for a cross-disciplinary view of AI research to maximize benefits and mitigate risks during integration into STEM disciplines. This report details the workshop's organization, format, participant demographics, outcomes, and conclusions, laying the groundwork for future collaborative efforts in this rapidly evolving field.

15. SUBJECT TERMS
Humans in Complex Systems; artificial intelligence; AI; machine learning; ML; Science, Technology, Engineering, and Mathematics; STEM; workshop; cross-disciplinary

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	44

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Steven Thurman	(520) 691-4791

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

Acknowledgments

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army or U.S. Government position, policy, or decision, unless so designated by other documentation. This research was supported by ARL, ARL contract number W911NF2410227.

1. Introduction

Recent advances in AI algorithms together with open-source access to foundational models,[1] large datasets, and compute resources[2] have created a rapidly expanding treasure trove of AI-ready resources for research. Simultaneously, novel applications of open-source models to scientific research and engineering problems have demonstrated promise in the integration of these tools as "crowd-sourcing" methods to enable the derivation of novel hypotheses. There are many challenges in both the fundamental research required to develop and explain trustworthy data-driven AI as well as the appropriate integration of these methods and tools into scientific research and engineering systems. To this end, initiatives such as the National Science Foundation (NSF) AI Research Institutes programs[3] provide funding pathways to stand up relevant organizational infrastructure. An important feature of the research institutes is the joint interdisciplinary nature of the AI research and the associated scientific research. However, different scientific disciplines are encountering overlapping problems in the application of new AI opportunities, particularly in the overwhelming amount of data available for processing. The Workshop on Cross-Disciplinary Challenges and Opportunities in AI Applications to Science and Engineering (C2OA2SE, pronounced "cozy") brought together a diverse set of speakers for a cross-disciplinary discussion of the common problems and AI approaches intrinsic in modern science and engineering. The objectives of the workshop were for participants to jointly

- Examine state-of-the-art AI approaches applied to science and engineering problems;
- Identify common features of best-in-class methods across disciplines;
- Share complex problems across our disciplines;
- Characterize classes of complex systems and approaches for rigorous investigation into these systems, including potential AI approaches to composition and decomposition, aggregation, modeling, entropy, and modularity; and
- Explore the integration of AI into said systems and examine what this tells us about engineering AI systems.

The organization of the workshop facilitated these objectives by combining technical talks, breakout, and group discussion sessions.

2. Organization

The workshop was organized by Dr. Sarah Kitchen, Deputy Director of Michigan Tech Research Institute (MTRI) in Ann Arbor, Michigan, in collaboration with Dr. Steven Thurman, Research Scientist with the U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL).

Conference Organizer: Dr. Kitchen received her Ph.D. in theoretical mathematics in 2010, with a research focus primarily on domains related to representation theory and category theory. Foundations required for these fields include the study of symmetries, transformations, decomposability of mathematical objects, dualities, and a wide range of complex problems that require a high degree of synergy between disparate topics. This background has provided Dr. Kitchen with the skills to read and interpret a wide range of theoretical and scientific literature and combine diverse areas of research into innovative systems and solutions. At MTRI, Dr. Kitchen has been researching problems related to intelligence, decision processes, and data representations for multiagent and distributed multisensor systems. Dr. Kitchen is keen to apply a synoptic view of AI research across science and engineering disciplines, which informed the planning, development, and organization of the C2OA2SE conference. Dr. Kitchen emphasized the need to bring together a variety of experts to prompt cross-disciplinary conversations about common themes, benefits, risks, and requirements for integration of modern AI into Science, Technology, Engineering, and Mathematics (STEM) disciplines.

Workshop Operations: Dr. Kitchen served as Program Chair for the first and third days of the workshop. Dr. Nancy French of MTRI served as Program Chair for the second day of the Workshop. Additional staff support included information technology and administrative services, refreshments, and scribes for the discussion sessions to facilitate effective and efficient reporting of outcomes of the event. A website was developed by MTRI staff prior to the workshop to provide scheduling and registration information.[*]

Venue: MTRI is a research center at Michigan Technological University and focuses on the research and development of signal processing, ML, and remote-sensing technologies for national security, environmental, and autonomy applications. MTRI's research staff consists of Ph.D. and M.S. holders in the fields of mathematics, statistics, physics, engineering, and computer science who have core competencies and research experience in statistical signal processing, information theory, ML, deep learning, robotics, remote-sensing technology, and sensor phenomenology. MTRI has an extensive, externally funded research

[*] https://www.mtu.edu/mtri/research/project-areas/machine-learning/c2oa2se-workshop/

portfolio funded by federal defense and civilian agencies, such as Defense Advanced Research Projects Agency, Air Force Research Laboratory, ARL, NASA, National Oceanic and Atmospheric Administration, National Institute of Standards and Technology, and U.S. Geological Survey, as well as state government agencies and commercial customers. Located in northeast Ann Arbor, Michigan, the MTRI facility is conveniently located near the Detroit Metropolitan Wayne County Airport and is walking distance from three hotels. Lectures for the proposed workshop were held in a large (60-person) conference/lecture room in the building where MTRI offices are located. The office space within MTRI contains conference rooms that were leveraged for breakout sessions. MTRI's proximity to the University of Michigan, as well as other state schools across the public university system (e.g., Michigan State, Eastern, and Wayne State), provided an additional educational opportunity by promoting the event to students.

Attendees: The C2OA2SE Workshop was an invitation-only event for speakers, with additional participation generated through limited advertising. There were 13 speakers at the Workshop, with 11 in-person and two virtual talks. There were 21 additional registered participants, 12 of whom were in person.

Speakers: The objective of the C2OA2SE was to prompt cross-disciplinary discussions about the use and role of AI in STEM fields. Figure 1 shows the successful diversity in fields of research by the speakers.

	Total Talks	Total Relations	Astronomy	Biomedical Sciences	Chemical Sciences	Complex Systems	Computer Science	Data Analysis	Earth Science	Life Sciences	Mathematics	Natural Language Processing	Science of Science	Scientific Computing	Statistical Machine Learning	Statistical Signal Processing
Astronomy	2	5					1	1						1	1	1
Biomedical Sciences	2	5			1		1		1	1		1				
Chemical Sciences	1	3		1			1					1				
Complex Systems	**3**	3						1			1		1			
Computer Science	**3**	7	1	1	1							1		1	1	1
Data Analysis	**3**	5	1			1					2				1	
Earth Science	**3**	4		1						1				1	1	
Life Sciences	1	2		1					1							
Mathematics	2	4				1		2							1	
Natural Language Processing	1	3		1	1		1									
Science of Science	2	1				1										
Scientific Computing	2	6	1				1		1						2	1
Statistical Machine Learning	**3**	**9**	1				1	1	1		1			2	1	1
Statistical Signal Processing,	1	4	1				1							1	1	
Totals	29	61	5	5	3	3	7	5	4	2	4	3	1	6	9	4

Figure 1. Fields of research of workshop participants, including speakers (13 talks) and overall attendees (21 registered attendees). Participants in multidisciplinary fields listed more than one field of research.

3. Program Format

The formal activities in the workshop consisted of talks, breakouts, and group discussions, which are defined as follows:

- **Talks:** Technical presentations by invited speakers with a 30-min talk and 15-min question and answer format.
- **Breakouts:** Small group discussions addressing specific technical questions, in a 1-h format with semistructured discussion templates.
- **Group Discussions:** Large group discussions developing insights and formulating conclusions, in a 1-h format with a mix of a semistructured discussion template and an open floor discussion.

The breakout session format is shown in Figure 2.

Breakout Topic 1

Time-Date	11:30-12:30 Thursday, 7 November 2024
Topic(s)	1. What research areas are still confronted with "too little data" in the era of "too much data"? How are current or emerging AI tools applied to each data regime (i.e. too little or too much)?
Moderator	TBD
Scribe(s)	T. Singleton
Objectives	• Identify similarities and differences across research fields, pertaining to the session topics • Define representative examples of similarities/differences • Propose actionable insights/next steps for rigorous inquiry
Format*	• Brief introductions (5 min) – name, organization, field of research, personal/professional interests in the Workshop, relevance of Session Topic to your field/research • Nomination of Facilitator/Moderator (if desired) (<1 min) • Review of Session Topic (5 min) – open floor suggestion for subtopics • Open discussion (35 min) • Robin take-aways pertaining to objectives (5 min) • Wrap up (10 min) – organization of Conclusions for Group Discussion * Times are rough/suggestions

Location: Kiva

Michigan Tech Research Institute

Figure 2. Example of a breakout topic.

The group discussion format is shown in Figure 3.

Group Discussion Wednesday, 6 November 2024

Date	Wednesday, 6 November 2024
Topic(s)	• Identify individual goals/objectives for the Workshop • What achievements of AI applications in STEM to date should be considered "breakthroughs"? • What lessons have we learned from these applications about the role of AI in STEM?
Moderator	S. Kitchen
Scribe(s)	M. Straub, T. Singleton
Objectives	• Identify similarities and differences across research fields, pertaining to the session topic • Define representative examples of similarities/differences • Propose actionable insights/next steps for rigorous inquiry
Format*	• Brief introductions (5 min) – name, organization, field of research, personal/professional interests in the Workshop, relevance of Discussion topic to your field/research • Review of Discussion Topic (5 min) • Open discussion (35 min) • Robin take-aways pertaining to objectives (5 min) • Wrap up (10 min) – Organization of Conclusions * Times are rough/suggestions

Location: Kiva

Michigan Tech Research Institute

Figure 3. Example of a group discussion topic.

4. Abstracts and Outcomes of Invited Talks

The abstracts presented in this section are derived directly from the authors who presented at the workshop and were a part of the workshop proceedings shared with attendees. The abstracts have not been edited from their original version.

4.1 Wednesday 6 November 2024

1300-1345	**Sarah Kitchen, Michigan Tech Research Institute**	**Welcome to C2OA2SE and Invitation to Interdisciplinarity**
Abstract: Modern artificial intelligence capabilities are both technically presenting us with new possibilities in our approach to research, as well as visions of what "could be" in the future capabilities of computing, from multiple perspectives, including algorithms, data analytics and high-dimensional statistics, large scale computing, data storage, processing and transfer, and applications. With the dramatic surge in applications of AI to scientific and engineering research over the last several years, we reflect on where we've come from and what we may need to integrate AI fully into research processes. Collaborations between AI practitioners and researchers across a wide variety of disciplines have proven effective in generating new discoveries. As we aim to learn from each other about how AI has played a role in research within our respective disciplines, we also consider what opportunities may lay ahead for new interdisciplinary studies.		
Outcomes: This talk prompted a discussion on how our understanding of human intelligence is flawed and the ways in which this provides an obstruction to interdisciplinarity.		
1345-1430	**Ryan McGranaghan, NASA-JPL**	**Science Understanding from Data Science (SUDS): Tying data science to physical science discovery and towards a multi-institutional movement**
Abstract: Physical science can greatly benefit from data science to overcome a variety of real challenges that are with us today. Data science isn't new, but large, complex data understanding has propelled it to the forefront of physical science. The needed data science for understanding of physical processes and systems in Astrophysics, Planetary Science, Earth Science, and Heliophysics is distinct from traditional data science efforts, such as automation and autonomy. Three key data science technologies have the ability to realize the needed capabilities. • Through physical model-based inference, we transform observation into understanding. • Through Uncertainty Quantification, we provide the rigor to draw valid conclusions. • Through Machine Learning, we gain computational efficiency, rapid discovery, and insight-building identification of key drivers. JPL's Science Understanding from Data Science (SUDS) endeavor brings these three key data science technologies together with physical scientists in a collaborative community that co-owns science inquiry. SUDS has created a new, vibrant community of practice where physical and data scientists work together to increase the speed, depth, and rigor of scientific return by revealing new connections through data science (including AI). This talk will discuss the JPL SUDS initiative, including the first SUDS Workshop and Conference held in August 2024.* We will conclude with findings from the initiative and workshop that reveal the need for SUDS to become a multi-institutional movement.		
Outcomes: SUDS had similar themes to C2OA2SE, with a more direct focus on data science applications to planetary and space sciences. Echoing the discussion of the first talk, a consensus at SUDS was that data science applications to science understanding have		

*https://sudsconf.com/program/workshop.html inconsistent project outcomes, which was attributed to human bias and the intrinsic challenge in quantifying how much we know about what we do not know. SUDS demonstrated that hands-on collaborative efforts are a pathway to overcoming the challenge in integrating these methods into the practice of science in these fields, but this is resisted culturally and institutionally.

1445-1530	**Bethany Lusch, Argonne National Laboratory**	**AI for Science and Engineering at the Supercomputing Scale**

Abstract: Researchers now have access to exascale supercomputers, which are primarily powered by GPUs. I will talk about how people are using supercomputers to apply AI to science and engineering. I will also discuss open challenges at that scale, especially when coupling AI and simulations. I hope to inspire you to think big!

Outcomes: Distributed computing is an implicit practice of high-performance computing. Critical gains in computing speed are achieved by deliberate decisions as to how to distribute complex calculations. For many applications, distributing via subsetting the data being processed suffices; however, with the advent of large foundation models, an important continuing area of open research is in methods for distributing the model itself. This is a problem with certain inference applications, not just a problem with training. In the case of training models that use large simulations, additional challenges arise with managing the data that is being produced as the models train. These problems will only become more important as we integrate AI methods with simulation methods, which is certainly of interest as an approach to generating representative synthetic data for training high-performance models for which the accumulation of real representative data is difficult.

1530-1615	**Paul Bendich, Geometric Data Analytics and Duke University**	**Shape-based Mathematics at the service of Deep Learning, with applications to AI Safety and Efficient Reinforcement Learning**

Abstract: Areas of mathematics that think about the shape of objects, in particular topology and geometry, are increasingly finding key application within the theory and practice of modern machine learning. This talk will survey two recent examples of this phenomenon involving work by the small Durham-based company GDA, and discuss many opportunities for future potential collaboration.

First, for safety and robustness of AI systems, we introduce topological parallax as a theoretical and computational tool that compares a trained model to a reference dataset to determine whether they have similar multiscale geometric structure. Theoretical and empirical evidence shows that model/data pairs which satisfy this criterion have highly desirable properties. Second, we introduce a reinforcement learning (RL) method that uses geometry to decompose sensor control problems and produce RL learning algorithms that produce high-performing sensor control policies that scale well as the number of sensors increase.

Outcomes: Topology and geometry tell us what it means to “preserve structure” as we create low-dimensional projections. Since data-driven AI models are produced by their training data, it is exceptionally difficult to define specifications that are meaningful for verification and validation of these models without independent methods for understanding the “shape” of the data on which they were trained. In classic statistical ML, the generalizability of the model is the hallmark of proper “fitting” of a model because it balances the bias-variance trade-off. In deep learning, the proper alternative to this concept is a method for validating that the “shape” of the model agrees with the “shape” of the data itself. This is a critical open area of research in AI.

4.2 Thursday 7 November 2024

0900-0945	**Carl Edwards, University of Illinois Urbana-Champaign**	**NLP for Scientific Discovery in Chemistry and Biomedicine**

Abstract: The world faces unprecedented challenges in the coming decades across climate change, healthcare, and food security, each requiring innovative scientific solutions that are scalable, adaptable, and cost-effective. Recent advancements in scientific NLP and multimodal AI—especially in integrating molecular data with natural language—present new opportunities to address these complex issues. A fundamental question, however, is why natural language integration with molecular modalities is crucial. This approach promises to accelerate scientific discovery by enabling 1) Generative Modeling for novel compound or hypothesis generation, 2) Bridging Modalities for addressing data scarcity, 3) Enhanced Domain Understanding in AI systems, 4) Automation of laboratory experiments and simulations, and 5) Democratization of scientific tools. This talk will explore these transformative trends and their potential to reshape science and medicine.

Outcomes: The most important open question regarding "AI for Science" is the right way to incorporate AI with human knowledge and tools, such as laboratory and simulation experiments, to create a robust system for discovery. Language model hallucinations are a strength in search applications, such as chemical discovery, but this needs to be combined with methods of validation. Training the language models themselves is a challenge, not just for molecules but for science generally, even if the models are trained with scientific texts, because what we understand about the world via science does change in time. Multimodality is going to be a key enabler to accurately connect scientific properties to natural language descriptions. New opportunities for collaboration across disciplines are possible through natural language descriptions that can identify and communicate field-specific scientific insights into language for nonspecialists.

0945-1030	**Steve Techtmann, Michigan Technological University**	**Applications of Machine Learning and Microbial Communities for Environmental Monitoring**

Abstract: Microbial communities are complex interacting systems that can sensitively respond to changes in the environment. Advances in genome sequencing has catalyzed an increase in data related to the composition and function of natural and engineered microbial communities. Previous work has shown that machine learning models trained on microbial community data can be used to address problems such identifying disease state and determining object provenance for source authentication. Here we leverage machine learning to use natural microbial communities as biosensors of environmental phenomena to determine the presence of pollution and ecosystem health. Our results indicate random forests models are very accurate in detecting the presence of crude oil and classifying the type of oil contamination in aquatic systems. Furthermore, models can be trained on as few as ten microbial features for prediction of oil contamination. However, these models suffer from low generalizability on other datasets from diverse locations. This lack in generalizability could be due to biological and experimental differences in datasets or due to limited training data. Future work is investigating methods for improving generalizability of models through identifying key steps in the wet lab processing of samples as well as the use of generative approaches to increase data for model training.

Outcomes: ML methods for classification are ubiquitous in scientific practice, and in the examples provided in this talk, they demonstrate similarities between the study of interacting and complex systems in biology with sensing system problems encountered in engineering. The primary distinction between the two is that in engineered systems, we generally have more control over the calibration of the sensing system itself, and in biological systems, we may be simultaneously measuring properties of the system as we solve our inference problems related to the system. The direct comparability of different systems and experimental practices may be ambiguous or unknown, and each independently varying factor can make the generation of sufficient data for discovering generalizable relationships among variables challenging. Disentangling the influence of process, environment, and data limitations on generalizability

results is necessary, but it is limited by the scalability of the process with respect to time, space, funding, and personnel. Generalizability in the usual sense of statistical ML means robust performance on new data drawn from the same distribution as the training data; however, in practice, we are often sampling new datasets from distributions that may not be the same as that of the training dataset.		
1045-1130	**Felix Herrmann, Georgia Institute of Technology**	**Digital Twins in the era of generative AI — Application to Geological CO_2 Storage**
Abstract: As a society, we are faced with important challenges to combat climate change. Geological Carbon Storage, during which gigatons of super-critical CO_2 are stored underground, is arguably the only scalable net-negative CO 2 - emission technology that is available. Recent advances in generative AI offer unique opportunities—especially in the context of Digital Twins for subsurface CO 2 -storage monitoring, decision making, and control—to help scale this technology, optimize its operations, lower its costs, and reduce its risks, so assurances can be made whether storage projects proceed as expected and whether CO 2 remains underground. During this talk, it is shown how techniques from Simulation-Based Inference and Ensemble Bayesian Filtering can be extended to establish probabilistic baselines and assimilate multimodal data for problems challenged by large degrees of freedom, nonlinear multiphysics, and computationally expensive to evaluate simulations. Key concepts that will be reviewed include neural Wave-Based Inference with Amortized Uncertainty Quantification and physics-based Summary Statistics, Ensemble Bayesian Filtering with Conditional Neural Networks, and learned multiphysics inversion with Differentiable Programming.		
Outcomes: Estimating the posterior distribution is the holy grail of inverse problems. Simulation-based inference allows one to apply ML to this process via conditional diffusion models. Multimodal data have traditionally presented significant challenges in the inverse problems community, but new ML methods are an enabler to their resolution. Namely, by correcting simulations using observed time-lapse data from the field, we can then condition our simulation on this observed data. The effort to achieve this system has been multi-disciplinary and actually integrates the data with ML. The ML methods are providing a new approach to extracting the probabilistic baseline and exercising probabilistic inference with much less compute required relative to traditional partial differential equation (PDE) techniques. Specifically, high-dimensional distribution estimation together with dynamics created computational complexity that limits the scale at which these methods can be applied. ML has provided new opportunities for STEM by providing alternative methods that facilitate scaling of the methods we have. However, as always, curation of appropriate training sets is a major challenge to our ability to apply these methods rigorously. The design of ML-enabled digital twins specifically requires an end-to-end vision of the requirements of the system for the intended applications of the twin, a principled way to detect and handle misspecifications, and calibration of the Uncertainty Quantification used in the model, which is still an open field of research.		
1330-1400	**Joshua Bloom, University of California, Berkeley**	**AI in Astrophysics**
Abstract: This talk digs into topics regarding the impact of AI in the landscape of astronomy research. The ability to apply AI to previously computationally expensive numerical processes enables the generation of incredible new images and data reduction applications. Simulation-based inference and surrogate models allow us to solve complex inverse problems.		
Outcomes: Astronomy has standard practices in effective ML-assisted "real-bogus" detectors that can separate real effects from spurious detections. The scale of the data in question is so large that researchers cannot save all the relevant data to a disk. Instead, the data runs through detection algorithms in memory, and the system saves the highest probability detections to the disk. Several problems in this area are related to the large volume of data versus the few labels for that data, which is a different version of the "small data" problem. Emerging techniques include multi-model, self-supervised models for clustering and anomaly detection, simulation-based inference, and neural density estimation for posterior estimation.		

1400-1445	**Santo Fortunato, Indiana University Bloomington**	**AI in Networks and the Science of Science**

Abstract: In this talk I will present some applications of AI in the fields of network science and science of science. Networks are everywhere. The graph representation of complex systems allows us to uncover how they work and how structure is related to function. A key problem of network science is community detection, i.e. the identification of groups of nodes which are more connected to nodes of their groups than to the others. Recently, embedding and clustering has been proposed as a new approach to find communities. We showed that some neural embeddings are surprisingly good at learning community structure. Another research question is whether deep neural networks have specific structural properties that are linked to their performance. We found that certain small motifs (balanced bicliques) are overrepresented in efficient networks. Also, neural networks have a near-optimal performance even when many of the edges are removed. Sparsification of neural networks may offer clues on how they operate. The science of science is the study of the structure and evolution of science via analysis and modeling of data on scientists and their interactions. Language embeddings are regularly used to map scientific papers and authors in low-dimensional vector spaces. The evolution of these maps over time helps us understand how science evolves, how innovation emerges and identifies the main ideas behind modern science.

Outcomes: Networks provide a flexible modeling approach to capturing relationships across entities from which we can apply graph theory to derive insights about those relationships. However, often bringing theory to practice is stymied by the computational complexity of the problems, such as community-finding. One approach is to leverage graph embeddings into a metric space to provide the opportunity to leverage proximity when associating nodes into communities. Structural properties of trained deep neural networks can be evaluated with these methods as well, and research is ongoing in the association of structure to function with these methods. New datasets and tools open the door for us to apply networks methods to the study of science as a system. This goes beyond the study of citation networks, and it now includes funding, mobility, collaboration dynamics, career dynamics, evolution of disciplines, and more. The content of full research papers can be now evaluated in context of the entire body of knowledge linked by prior and future papers linked through them. Historical analysis shows that science advances by explosions of breadth, followed by depth, not through incremental advancements within niche specializations. The evolution of the landscape of knowledge might allow for predictions of future innovations, and LLMs can help us uncover the backbone of science that would enable these predictions.

4.3 Friday 8 November 2024

0900-0945	**Raphaël Pestourie, Georgia Institute of Technology**	**Scientific Machine Learning for Optimization via Surrogate Models**

Abstract: This talk is about the machine-learning-enhanced optimization of scalar functions within the context of metamaterials—materials that derive their properties from their geometry—which lends itself to large-scale optimization. We will introduce surrogate models for scalar functions in the realm of scientific machine learning, where the output corresponds to a scalar property of a PDE solution. Next, we will explore how surrogate models differ from traditional solvers by showcasing two distinct methods that leverage multi-fidelity models: multi-fidelity DeepONets and Physics-Enhanced Deep Surrogates (PEDS). Using the illustrative example of optical metasurface design, we will demonstrate how surrogate-based approximate solvers can alleviate the simulation bottleneck and facilitate the large-scale optimization of complex systems. Finally, we will end with a speculative discussion on how machine learning techniques could enhance optimization by identifying the relevant degrees of freedom that drive physical responses and high performance.

Outcomes: The work of Dr. Pestourie provides an excellent example of complexity reduction in numerical methods for PDE solving via symmetry (periodicity) and low-dimensional surrogate models. The neural network approach used synthetic data generated by a physics-based simulator—this is an example of where the physical model is known formally, in general, and can be used to validate the data-driven model itself. An important point in the success of this work is the expert design and integration of the AI methods. The context of where this approach fits into the larger process was understood, the trade-off of complexity versus accuracy versus training time and others were all estimable before implementing the methodology.

0945-1030	**Allan Avila, AIMdyn, Inc.**	**Spectral Operator Methods for Learning Dynamical Systems**

Abstract: This talk covers two main developments: advancements in AI model analysis using Koopman operator theory (KOT) and the extension of KOT beyond function spaces. First, we discuss how KOT can enhance neural network pruning, classify algorithms, and provide insights into training dynamics across network architectures. Next, we discuss recent work extending KOT beyond the setting of function spaces which offers a spectral framework that captures additional system properties. This extension necessitates novel numerical methods for spectral approximation, which we discuss.

Outcomes: Spatial mode analysis discovers subfunctions and eigenvalues that superimpose to reconstruct the original data. These methods have been applied to identify when nodes can be pruned and when models and training routines are effectively equivalent, via the Koopman spectra. There are many interesting theoretical questions about how to generalize the Koopman Operator Theory to broader classes of systems than its original application (measure-preserving dynamical systems). Changes in class remove the theoretical guarantees of the method, but like Dr. Bendich's talk, it provides external-to-AI-practice, mathematical formalisms that could be used for building validation methods.

1045-1130	**David Roy, Michigan State University**	**Artificial Intelligence for high spatial resolution satellite fire monitoring**

Abstract: Geospatial artificial intelligence, an emerging discipline that combines innovations in data science and deep learning methods, has the potential to revolutionize scientific understanding of our changing planet and meet the outstanding demands for actionable information. Satellite data have a long provenance for fire monitoring. Fire products that map the occurrence and intensity of fires and the area burned have been provided by space agencies for the last several decades. The commercial sector plays an increasingly important role, and geospatial analytics and Earth Observation companies are providing satellite imagery and information services with extraordinary new capabilities. The evolution of satellite fire monitoring is presented using the August 2023 wildfire disaster over the island of Maui, Hawaii, as a case study. Recent results of a deep learning Swin-Unet model to map 3 m burned areas using commercial PlanetScope data are presented. The limitations of currently available space

agency satellite fire products and the needs and potential for future satellite monitoring and prediction of wildfire disaster events, are highlighted.

Outcomes: This talk prompted a discourse in the definition of digital twins, with the important note that digital twins include the ability to impose controls that perturb the modeled system, which is an enabler to making Monte Carlo projections. The primary problem defined in the talk was about the detection of small fires. We have ample Earth science data, but we cannot detect small fires as they happen, just evidence of burns after the fact. The temporal resolution—a faster revisit rate is needed to obtain early detections for fires. Physical climate models today integrate AI with ensemble forecasting methods, but we have no such model for fire dynamics, the role of fire in the carbon cycle, or regional fire prediction. The work presented demonstrates the role of AI in enhancing our ability to map burned areas, including through occlusions such as smoke, and therefore making more use of the data we have, and is another example of how expert knowledge in a field drives the appropriate integration of AI into a scientific application. Dr. Roy's talk emphasizes two emerging themes of the Workshop: 1) We can do more with the data we have by incorporating "AI methods" (taken broadly to include data science, computing, and statistical ML in addition to deep learning); and 2) Science suffers from a lack of data to solve the problems the experts want to solve, even in the era of "big data."

1130-1215	**Lingfei Wu, University of Pittsburgh**	**AI for Team Science**

Abstract: While traditional team science has focused on the benefits of teams, my study pioneers an examination of the costs of teams. I leverage big data and AI to quantify scientists' roles within teams, breaking new ground with analyses of self-reported author contribution data,[4] and explaining how large teams can impede individual creativity.[5,6] My team and I are now advancing to the next phase by creating and analyzing the world's first observational dataset on author contributions, drawn from author-specific macros in two million LaTeX source code on preprint websites. This approach enables us to explore how large teams may obscure individual recognition, contributing to the long-standing inquiry into the contribution-credit paradox in science and beyond, originally highlighted by the sociologist R. K. Merton in the 1960s.

Outcomes: Dr. Wu's heuristic interpretation of his results is that as teams grow larger and more dispersed, they linearly scale "muscle" researchers executing the process of science, and they only logarithmically scale "brain" researchers conceiving of the ideas for the research programs. Large teams succeed in division of labor more than those in innovation. The takeaway cost-benefit analysis of his work is that small teams disrupt (cost of size), flat teams innovate (cost of hierarchy), and on-site teams conceive (cost of distance). In relation to the call to interdisciplinarity, small, flat, local teams of researchers are likely to be the format needed to synthesize multiple fields of research into new innovations.

5. Breakout Sessions

The breakout topics discussed during the workshop were as follows:

1. What research areas are still confronted with "too little data" in the era of "too much data"? How are current or emerging AI tools applied to each data regime (i.e., too little or too much)?
2. How do we ensure data-driven AI methods meet scientific, engineering, and mathematical rigor?
3. Is there an inevitable emerging role for interdisciplinary research between computing/computer science and other STEM fields?
4. What are the challenges in training the next generation of scientists and engineers when combining "data-driven" methods with expert-driven practices?
5. What metrics adequately quantify uncertainty in AI output for STEM research purposes? What are the desired properties of such metrics?
6. What lessons have we learned about the ways in which current AI methods fail to meet STEM needs?
7. How are STEM fields leveraging AI to improve human understanding of complex and previously intractable research problems?
8. How can AI act as an enabler to interdisciplinary research?

5.1 Topic 1 Key Findings

Discussion Question: What research areas are still confronted with "too little data" in the era of "too much data"? How are current or emerging AI tools applied to each data regime (i.e., too little or too much)?

- New discoveries tend to occur in the "tails" of the distributions—finding the right place to look is the industry of science—these tails have the problem of presenting as "weak signals" in large amounts of data ("too much data"), and simultaneously will not have sufficient representative data within a dataset for purely data-driven models to robustly fit that area of the distribution. Drug discovery was used as a representative example.
- New AI tools are in the beginning stages of being applied to synthesis of data in low-data areas—there is a lot of interest in generating synthetic data for various applications, but AI applications are also being developed for automation of experiments to generate real data.
- New methods are needed to leverage the manifold hypothesis, including translation of results about model pruning and effective engineering of AI models, with additional understanding of the data being used to train the models.
- Active learning or human-in-the-loop tuning seems to be necessary to adapt performance of purely data-driven approaches to application in rigorous disciplines such as science and engineering.

- It is unclear whether this requires a complete paradigm shift or to what extent combinations of existing methods with various amounts of human expertise (e.g., feature engineering) are required.

5.2 Topic 2 Key Findings

Discussion Question: How do we ensure data-driven AI methods meet scientific, engineering, and mathematical rigor?

- Most of our techniques for engineering rigor start with an assumption of a model of a known type or within a class of standard models, in which case formal understanding of characteristics of the model class can be used for rigorous validation techniques. Data-driven models are fundamentally different, and they may require more of a natural sciences point of view; however, since they are engineered systems, they may require a "meet-in-the-middle" approach.
- The main challenge in translating rigor from science and engineering to data-driven AI methods is that these approaches tend to require a mathematically or statistically rigorous and justified model as part of the process. Mathematical and statistical rigor include proof-based methods for variable selection, model reduction, and error and asymptotic bounding of the model. This includes techniques such as local linearization and fitting previously established models based on axioms and possibly simplifying assumptions about the underlying system processes. Different fields have different standard practices for simplification based on the theory of the field. In application, these models are approximations of natural and man-made systems, developed over time with human insight and judgment in addition to the scientific method; therefore, sensitivity analysis for the model is a necessary part of the developmental process, to establish quantifiable bounds on "how badly" the model can fail when it does fail. Data-driven models at this time have no such accepted methodologies for analogous sensitivity analysis, despite several industry-accepted performance metrics, because, as universal function approximators, large models trained on large datasets can be assumed to only match the quality of the training data.
- Since techniques such as uniformization of data can assist in improving these models, as well as fine-tuning with human feedback, we will not be able to fundamentally separate rigorous analysis of these models from rigorous analysis of the training data.
- Based on the standard techniques from engineering applications, local linearization is a primary tool. If this is not possible, statistical methods

should be applied and may require data generation or augmentation of some form to test the hypotheses about the model.

- "Reliability" may be an underexplored area of interrogation with respect to AI models, namely, model results must be shown to be consistent, even when they are not deterministic.
- Uncertainty quantification is a known research area, but it includes the problems of calibration and definition of metrics and identification of nuisance parameters.

5.3 Topic 3 Key Findings

Discussion Question: Is there an inevitable emerging role for interdisciplinary research between computing/computer science and other STEM fields?

- Science itself is heading toward interdisciplinary research, but University infrastructure and funding mechanisms are lagging behind the needs of researchers in producing this work. Data availability permeates all disciplines, so reviewers of research proposals, tenure applications, and journal articles need to be adequately "data science/ML" aware to evaluate modern research in these fields. Either an unusual degree of breadth is required for an individual or multiple, specialized people are needed for the co-review process. The challenge then becomes effective communication across those disciplines.
- Incentive structures require a change to support the emergence of necessary interdisciplinary research. Few programs fund interdisciplinary research explicitly. Academic departments are assumed to be siloed, and there is a reluctance to fund outside departments. Co-principal investigator (PI) structure of cross-disciplinary proposals misalign the needs of tenure-track faculty (sole PI grants) with the needs of the research community (collaborations).
- NSF initiatives combining AI with other disciplines, as exemplified in past NSF dear colleague letters, sometimes target both fields (NSF research institutes), but sometimes they do not.
- Grant panels may be unbalanced toward one discipline, and this is a blind review, so it may be challenging for a truly interdisciplinary team to stand out. Likewise, in the produced research, attribution to researchers of quality work by different communities may be difficult to communicate to tenure-track committees, which is a barrier to young researchers who want to make an impact with interdisciplinary work.
- At this time, it is unclear where the responsibility for "AI" belongs at a university, what the relationships are among roles for AI across

departments, and how that serves the students in terms of workforce preparedness.

- The community of people who can evaluate the impact of interdisciplinary work will be the drivers of that research. Communication across specialized disciplines will be necessary, but not for computing. AI can be an enabler to communication.

5.4 Topic 4 Key Findings

Discussion Question: What are the challenges in training the next generation of scientists and engineers when combining "data-driven" methods with expert-driven practices?

- Pedagogical decisions must be made, given the finite time per semester—how do we divide the time between formal analytical and the "AI" alternative? The analytic approach is important for understanding the foundations of the theory, but it may not be as impactful from the perspective of workforce development
- Using AI and teaching AI can increase both work placement and salary increases, so proficiency in these methods is of unquestionable value.
- Creativity and full understanding of a subject comes from direct experience with fundamental rules and principles—this must be preserved, even throughout the integration of AI tools. Development of solutions to complex problems requires knowledge of what problems other experts can and cannot solve. Co-curricular development or co-teaching allows you to accumulate this knowledge.
- Without understanding human intelligence, we cannot understand the full impact of integration of AI into any human system. AI experts need additional knowledge about human decision-making to provide adequate suggestions about the role, integration, and appropriate influence of AI.
- How to use AI in support of STEM practices is more important for most coursework than complicated model development, but the AI developer may not be able to produce an appropriate model for a specific task without deep understanding of the use case.
- Experiential and/or interdisciplinary capstone coursework may present an opportunity to synthesize fundamental knowledge gained over an undergraduate curriculum, taken together with practicable, hands-on experience for using AI models to analyze field-specific problems.

5.5 Topic 5 Key Findings

Discussion Question: What metrics adequately quantify uncertainty in AI output for STEM research purposes? What are the desired properties of such metrics?

- For STEM purposes, each research community will need to define their own metrics relevant to the models, applications of the models, and associated data involved in the research problems for that community. It is unlikely that general metrics will fully capture the needs of a research community. This may be one of the challenges in integrating AI into STEM in a rigorous fashion—as with peer review, the community of experts will need time to adapt general approaches emerging from industry to their specific applications appropriately.
- Expectations for AI are high relative to the actual practice of developing AI models—specifically, we want the models to generalize well and perform well with assurances on data that we do not have to test during development. The push for bigger models has subsumed the transfer learning problem, but the challenge remains.
- Different types of AI models and different applications have different notions of uncertainty and different requirements for quantification of uncertainty. The downstream application of the output of the model determines what the error tolerance is for the model output. Variability in the output (e.g., standard deviation) suffices in some cases; however, variability is not always uniform over the tasks a general model is developed to perform, and downstream applications must also be robust to that nonuniformity.
- Desired properties include the following: high accuracy and precision on test datasets; models relating these values to properties (including sample size) of the training data; and generalization error on novel test sets. However, inaccuracy in the training data also needs to be understood and quantified—this may be underrepresented in how we communicate about the capability and performance of trained models.
- This is an open area of research in AI, including explainability, interpretability, "safety" for neural networks, rigorous verification and validation methods, and metrics. Potentially, starting from quantifying "risk" for the system performance and working backward to "uncertainty" metrics for integrated AI models is appropriate, if the propagation of errors is what really matters—this is more relevant for engineering systems than for things such as hypothesis testing or classification problems.
- There is a decision to be made about whether to fundamentally treat the uncertainty in these models the same way we treat uncertainty in human

systems and relations. People are not rational, but we define trust and estimate uncertainty relative to human–human interactions based on prior interactions and mental models of intent, among others. Are we destined to be limited to the same with AI?

5.6 Topic 6 Key Findings

Discussion Question: What lessons have we learned about the ways in which current AI methods fail to meet STEM needs?

- The "desire" for AI applications is automating rote tasks and otherwise "making life easier," but the gap between the methods and these needs is the appropriate data to train the models and new methods to do "more" with less data. Benchmarks used to compare models are not necessarily relevant to applications in STEM.
- The data also needs to be clean and correct, because the models will be limited by data accuracy—reusing models trained on bad data will propagate this throughout the community. A significant need for STEM applications of AI methods is validated, assured, and robust datasets that meet specified standards of quality, so that models trained on these data are built on data that meets an expert-agreed definition of "quality" for STEM purposes.
- The speed of iteration for model development and innovation is out of synchronization with the time and resources required to generate relevant data for these problems or the relevant science. AI researchers at the forefront of their fields will not be developing techniques to address specific problems, and the researchers are oriented toward datasets that are not necessarily relevant to significant STEM problems. STEM researchers are almost by necessity going to be using out-of-date models and techniques.
- Data-driven AI is always going to perform the best in instances in which there is sufficient quality data, but STEM needs include rare-event prediction and performance quantification. AI is often presented as "a" solution, where its best use may be considered as "part of" a solution.
- The robustness and generalizable performance of large models is too energy and compute intensive, which limits applicability and usability, particularly in edge systems.

5.7 Topic 7 Key Findings

Discussion Question: How are STEM fields leveraging AI to improve human understanding of complex and previously intractable research problems?

- The talks at the workshop have the common theme that AI has the capability to explore large, complex spaces beyond the capability of humans to hold all of the complexity in memory. Attention mechanisms have demonstrated use cases in identifying for humans the interesting patterns in a problem space that warrant active inquiry. This could possibly be transitioned to searching "ideas space" and to capture salient aspects of data or connections across fields.
- STEM is invested in breaking complex problems into more understandable, smaller problems, and many fields are designed to focus specifically on niche elements. AI has the potential to facilitate communication across these niche fields by enabling translation of results and techniques.
- Science has a known problem with an explosion of research literature but a decrease in innovation relative to that output. AI can now additionally be leveraged in science for massive data triage, document generation, search, and process automation, which replaces some human functionality; however, on its own, AI does not lead to insights.
- Potential ideas for AI applications, possibly at the level of scientific corpus evaluation and summarization, are as follows: identification of assumptions, limiting processes, self-isolating research communities, or other human behaviors that produce dead-end STEM work; automatic translation and/or identification of overlapping concepts across research fields, where multiple perspectives could lead to new results; analysis of reproducibility of research; summarization of fields of research; and identification of emerging standards of practice from new and rapidly evolving fields (such as AI itself).
- Benchmarking in AI is problematic—we have an explosion of benchmarking datasets and metrics, each resulting in a flurry of activity, which has an increasing volume of repurposing previous methods with limited new ideas and approaches. "Repurposing" and assembling systems is a much faster process than developing something truly new.

5.8 Topic 8 Key Findings

Discussion Question: How can AI act as an enabler to interdisciplinary research?

- Simulation-based inference is a natural place for AI to enable interdisciplinary research, because each discipline can bring its own appropriate models and laws to the appropriate part of a joint simulation.
- The general impression in STEM is that AI researchers are opportunistically using datasets to publicize capabilities that could be even further improved if they had worked with domain experts, but it is unclear who will fund those collaborations.
- AI will enable interdisciplinary research between domain experts who want to choose a model "off the shelf" and integrate it with other components relevant to their research problems. The usefulness of a data scientist without a specialization in a STEM discipline may be limited.
- Conversely, AI researchers are developing foundation models that can provide value to researchers through fine tuning. If they incorporate domain experts in their development of the models, they may produce more useful output or more sample-efficient models. This is likely an under-researched problem in general, although it presents an obvious opportunity for collaboration between AI researchers and individuals in STEM disciplines. The embedding of physical laws, symmetries, or other bounds and constraints is not well understood in AI in general, although this is an active area of research.
- LLMs can be used as interdisciplinary translation "copilots," where a researcher can query basic information and jargon from another domain to rapidly gain context—if they are also working with an expert in that domain, and the output is a summary or only partially correct, it is not a substantial hindrance. Applications of this type that can accurately translate results into interoperable standards could be a massive enabler to interdisciplinary research.

6. Group Discussions

6.1 Wednesday Group Discussion

Discussion Question: What achievements of AI applications in STEM to date should be considered "breakthroughs"?

- AlphaFold was the unique example identified as a scientific breakthrough enabled by AI. Large language models (LLMs) have opened the door to translating many sequential prediction problems across STEM into these types of models.

- LLMs are also enabling a great deal of rapid capability development in task generalization in robotics, as well as natural language interfaces for human–AI partnership. Artificial agent generation through imitation learning methods are an emerging capability
- Brain–computer interfaces are not examples of brand new technology; however, recently, there has been substantial advancement through analysis of large datasets with annotated brain events.
- AlphaFold and advances in brain–computer interfaces demonstrate the requirement for clear validation criteria for correctness in the output of the models.
- Open questions in human–AI teaming and interfacing were cited in both directions: we are making a lot of progress in training AI through human feedback, and we are improving brain–computer interfaces through ML applied to large datasets.
- Common challenges to understanding AI and interdisciplinarity may be due to limitations in understanding human cognition, theory of mind, and more.
- Common themes in speaker and participant interests were the mismatch between desires for AI and current capabilities, integration of simulation and digital twins with AI as a scientific and engineering enabler, and the role of supercomputing as a facilitator.

6.2 Thursday Group Discussion

Conclusions from Thursday Breakouts

- In general, there is too little data, especially in science fields, and especially for the frontier questions, that should to be explored by scientists.
- The effort needed to feature engineer or otherwise design your model depends on the data you have.
- Data-driven AI methods can only meet scientific, engineering, mathematical, and statistical rigor when they are explainable and outcomes are reproducible.
- The main needs for interdisciplinary research are funding sources and support from academic institutions to incentivize this type of work. Technically, methods to improve effective communication across disciplines are needed to lower the transaction cost of embarking on this type of work. There was a consensus that there is a great need in STEM for truly interdisciplinary work, which may include AI researchers, but they do not displace domain experts.

- A “middle path” is needed in educating the next generation of our workforce that combines both analytical practice and rigor with data-driven methods. In addition, the sparse data problems in STEM will never disappear, so a thorough understanding of the diverse array of statistical data science methods available to overcome data limitations continue to be substantially important for workforce education.
- No consensus was reached on metrics that can adequately quantify uncertainty, but framing first from the perspective of risks may be useful for extracting appropriate metrics for individual applications and fields of research. AI industry standards of train/test splits are not necessarily the correct approach.
- Consensus was that the group have been confronting the same general questions across their diverse disciplines and research areas. This implies that we are addressing the correct questions in the workshop.

6.3 Friday Group Discussion

Conclusions from Friday Breakouts and Workshop Overall

- AI is marketed as “the solution” to a problem rather than a part of the solution.
- The interface between the model and the user (or user problem) is a critical component for developing the correct model for a problem and associated metrics for performance.
- Related to this is the need to refocus on methods for lower-energy models, which have not been the focus of the highly publicized AI community in recent years.
- Breakthroughs in AI and STEM domains may remain staggered, although the rate of change in AI is high, so it is unclear how to most effectively translate capability to STEM problems.
- Benchmarks used in AI research are not necessarily useful or meaningful for applications in STEM. General principles for the types of data and problems with different methods and architectures are useful for the community well, but specific metrics for performance, quality, uncertainty quantification, and others will be domain specific in STEM applications. Research communities may have to be responsible for defining standards of practice for integration of these methods into their fields.
- AI can enable interdisciplinary research by facilitating communication as an “expert copilot” from another domain. Language models are well developed enough to provide this as a service with low cost to entry and otherwise facilitate communication across teams.

- AI may play a role in a cyber-physical system version of citizen science, by providing substantial detection and classification capabilities on small, distributed, embedded sensors.
- AI can facilitate science research by encoding the full complexity of a dataset strictly beyond human comprehension, then providing that model as an embedded representation for further inquiry.
- In the future, researchers in STEM fields need to establish standards, methods, and ethics around data, data availability, and data quality.
- Digital twinning is an emerging methodology that combines modeling, simulation, data streams, and AI as a "living" representation of a system. Uncertainty quantification in these models is a significant open problem, and it is an area that is currently receiving funding for research.
- Computational resources, time, funding sources, and academic incentives are not quite aligned for extracting the most value from AI applications to STEM.
- The STEM workforce is changing, and AI will be a contributor to this change.

7. Outcomes

This section summarizes the outcomes of the discussions as they pertain to the original objectives of the workshop.

7.1 State-of-the-Art AI Approaches Applied to Science and Engineering Problems

The presentations over the course of the workshop demonstrated a diverse array of AI methodologies applied to different STEM disciplines. Not all approaches were strictly applications of neural networks models, but where those models were used, they were integrated as specific components of more rigorous frameworks for inquiry. The design of the broader model or context for the application of the AI method required domain-specific insight, and this is expected to continue. LLMs and similar transformers used to develop foundation models are some of the newest methods in AI, and they are currently receiving the most attention by broad audiences. In addition, deep learning and transformer models have simultaneously been used to achieve breakthroughs such as AlphaFold.[7]

7.2 Identify Common Features of Best-in-Class Methods across Disciplines

The common feature across disciplines was not specifically related to AI models but to the availability of data that is necessary to train the models, and more specifically data that is of a sufficient quality for training. In some applications, reliable data generation or synthesis is a straightforward part of the process. These settings are usually associated with engineering-type modeling and simulation, where standard practices are well established. In science applications, data generation is not always as simple, but established practices and standards exist for data collection and validation from traditional methods, where the new opportunity provided by AI is in scaling and uniformizing these studies, as well as aggregating methods and data streams into synoptic models such as digital twins.

7.3 Share Complex Problems across our Disciplines

The speakers shared complex problems associated with their research disciplines. Through these talks and discussion sessions, complexity became understood as a challenge with the boundaries of properties at scale—either the scale of the system or the scale of the data. In both cases, high dimensionality provides challenges to the discovery and modeling of interrelationships among the variables. Mathematical fields such as graph theory, topology, geometry, and dynamical systems have theoretical and numerical/computational methods to begin exploring the structure of these relationships, even in latent spaces produced by AI models, and this is an opportunity for new mathematical methods to be introduced to AI research.

7.4 Characterize Classes of Complex Systems and Approaches for Rigorous Investigation into These Systems

The characterization of complex systems and decomposition approaches were not explicitly addressed through the breakouts and discussion sessions, although questions about rigor in modeling were discussed, as well as the requirements for both science and AI to contribute effectively to digital twins. A significant problem in complex systems is the association of metrics and behaviors across scales. Analogously, large, intractable problems in STEM fields include the same type of "decomposition" problem, that is, the community researching the problem isolates subproblems and develops specialized methods to solve those problems, without requiring the larger context of the original problem. We are introduced to and trained in specific STEM disciplines that have been developed after this decomposition has already occurred. The practices we individually learn in our professional education may be constrained by 1) the cultural norms of the field as well as 2) the availability of mentors, collaborators, and/or funding. The scale at which unsupervised and semisupervised learning methods in AI are now reliably applied to specific problems within STEM disciplines and workflows has provided new opportunities to scale the process of science itself. Expertise in a field leading to a successful research career is about asking the right fundamental questions to which that process is applied. If the intractability of complex problems in STEM really occurs across larger communities consisting of multiple subfields or disciplines, the real opportunity may be in identifying these new fields or in generating recombinations for new ideas and techniques to be developed. This could include "meta"-fields within the discipline that are intrinsically multidisciplinary.

7.5 Explore the Integration of AI into Complex Systems, and What This Tells Us about Engineering AI Systems

This question appears to be at the forefront of inquiry into the best use of AI. Digital twins perhaps provide a test case for the appropriate integration of AI into a system that will reveal problems and lessons to be learned about engineering AI systems. From the practices of modeling and simulation and systems engineering, as well as the National Academies report on digital twins,[8] we know that a fundamental challenge in digital twins is uncertainty quantification. This is also a fundamental challenge in data-driven AI. Simulations based on known dynamical systems can be verified rigorously because we know from formal analysis what the properties of the correct output for the simulation should be. Therefore, when developing complex simulations of engineered systems from component models that have been verified to be correct and for which we understand error in the models, we are able to project the propagation of these errors through the entire system. At the same time, digital twins provide the opportunity to integrate real data streams, as well as conventional modeling, simulation, and statistical analyses of the data, which may be used to provide a system of "checks and balances" to AI models integrated into the larger whole.

8. Conclusions

The emerging role of AI in STEM is still in an exploratory phase, in which the best understood models and methods are being integrated by domain experts in targeted ways. The main limitation of AI integration into STEM is in the availability of data of an appropriate, verified quality for scientific rigor. This has always been a time-intensive task for science that is not intrinsically remediated by data-driven AI models, although the potential value of generative AI in the data-generation process is a source of excitement around these methods. The integration of multiple models into larger scientific systems, such as digital twins, could provide the next frontier for data availability and scientific model development and validation, through combinations of real-time data feeds and simulations with scientifically rigorous models. A synoptic interdisciplinary view of the complex integration of scientific research paradigms may enable the specification of desiderata for individual models that are not yet known. One significant requirement for advancing work with this approach is interdisciplinary collaboration across research fields and/or subfields. Culturally and institutionally, this presents a challenge, including in the professional development and education of the next generations of researchers. The current generation of emerging Ph.Ds. who have grown up with interests spanning computing and science are best positioned to collaborate among subject matter experts in specific disciplines and the AI, computing, and data science communities. Traditionally, funding and the tenure-track academic pipeline has been challenging for interdisciplinary researchers to navigate, due to the difficulty subject matter experts have in evaluating interdisciplinary work individually. These challenges can be addressed by increased awareness by funding and professional communities. A related challenge identified by team science research is that the format needed to synthesize multiple fields into new innovations has traditionally been small, flat, and local, which is increasingly unlike research team structures. Increasing awareness of this conflict, the STEM and AI research communities can correspondingly focus on the way we work sociologically as we adopt these new tools.

9. References

1. Hugging Face [homepage]. Hugging Face, Inc; c2025. https://huggingface.co/models

2. NAIRR Pilot. Advancing US innovation in artificial intelligence. National Artificial Intelligence Research Pilot (US); 2024 Jan. https://nairrpilot.org/

3. National Artificial Intelligence Research Institutes. National Science Foundation (US); 2023 July 31. https://new.nsf.gov/funding/opportunities/national-artificial-intelligence-research

4. Xu F, Wu L, Evans J. Flat teams drive scientific innovation. Proceedings of the National Academy of Sciences. 2022;119(23):e2200927119.

5. Wu L, Wang D, Evans JA. Large teams develop and small teams disrupt science and technology. Nature. 2019;566(7744):378–382.

6. Lin Y, Frey CB, Wu L. Remote collaboration fuses fewer breakthrough ideas. Nature. 2023;623(7989):987–991.

7. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–589.

8. National Academies of Sciences, Engineering, and Medicine. Foundational research gaps and future directions for digital twins. National Academies Press; 2024. https://doi.org/10.17226/26894

Appendix A. Speaker Biographies

Allan Avila earned his B.S. in Mechanical Engineering from the University of California, Riverside, in 2015, followed by an M.S. and Ph.D. in Mechanical Engineering from the University of California, Santa Barbara, in 2017 and 2020, respectively. He is a member of the Society of Industrial and Applied Mathematics and a recipient of the 2018 Hispanic Scholarship Award. Currently, Allan is a senior researcher at AIMDyn Inc., where his research focuses on data-driven analysis and forecasting of dynamical systems through spectral operator methods.

Paul Bendich is Chief Scientist at GDA and is a Research Professor of Mathematics at Duke University. He received his Ph.D. in Mathematics from Duke in 2008 and held postdoctoral positions at the Institute for Science and Technology Austria and Penn State. Dr. Bendich's doctoral work laid some of the early theoretical foundations for topological data analysis (TDA). Since then, he has been at the forefront of the integration of TDA with more standard ML and statistical techniques. This work has wide application in vehicle tracking, brain imaging, and image simplification, among many other areas.

Joshua Bloom is an Astronomy Professor at the University of California, Berkeley, where he teaches radiative processes, high-energy astrophysics, and a graduate-level "Python for Data Science" course. He has published over 300 refereed articles mainly on time-domain transient events, AI, and telescope/insight automation. His book on gamma ray bursts, a technical introduction for physical scientists, was published by Princeton University Press. Josh has been awarded the Data-Driven Discovery prize from the Gordon and Betty Moore Foundation and the Pierce Prize from the American Astronomical Society; he is also a former Sloan Fellow, Junior Fellow at the Harvard Society, and Hertz Foundation Fellow. He holds a Ph.D. from Caltech and degrees from Harvard College (A.B.) and Cambridge University (MPhil). He was co-founder and Chief Technology Officer of Wise.io, an AI application startup, acquired by GE in 2016.

Carl Edwards is a Ph.D. candidate in computer science at the University of Illinois Urbana-Champaign where he is advised by Professor Heng Ji. In terms of research interests, he is generally interested in models that leverage multiple modalities and domains of data centered around language. His research interests cover information extraction, information retrieval, natural language processing, representation learning, text mining, and transfer learning. In particular, his work applies these tools (and develop new ones) to scientific texts to accelerate scientific discovery. He is currently working on projects related to chemistry literature in association with the Molecule Maker Lab Institute, an AI Institute supported by the National Science Foundation (NSF). In general, he is passionate about language and multimodal for compositional function-level control of drugs, proteins, and material design.

Santo Fortunato is a Professor at Luddy School of Informatics, Computing, and Engineering of Indiana University. Previously, he was Professor of Complex Systems at the Department of Computer Science of Aalto University, Finland. Professor Fortunato got his Ph.D. in Theoretical Particle Physics at the University of Bielefeld in Germany. His focus areas are network science, especially community detection in graphs, computational social science, and science of science. His research has been published in leading journals, including Nature, Science, Nature Physics, Proceedings of the National Academy of Sciences (PNAS), Physical Review Letters, Physical Review X, Reviews of Modern Physics, and Physics Reports, and he has collected over 48,000 citations (Google Scholar). His single-author[1] article is one of the best-known and most cited papers in network science. Fortunato received the Young Scientist Award for Socio- and Econophysics 2011, a prize given by the German Physical Society, for his outstanding contributions to the physics of social systems. He is a Fellow of the Network Science Society (2022) and the American Physical Society (2022). He is the Founding Chair of the International Conference of Computational Social Science (IC2S2), which he first organized in Helsinki in June 2015. He was Chair of Networks 2021, the largest ever event on network science, a historical merger of the NetSci and Sunbelt conferences. He is the author of the book titled *A First Course in Network Science*,[2] the most accessible textbook on the new science of networks.

Felix J. Herrmann is a Professor with appointments at the Earth & Atmospheric Sciences, Computational Science & Engineering, and Electrical & Computer Engineering at the Georgia Institute of Technology. He leads the Seismic Laboratory for Imaging and Modeling (SLIM) and he is co-founder/director of the Center for Machine Learning for Seismic (ML4Seismic). This Center is designed to foster industrial research partnerships and drive innovations in AI-assisted seismic imaging, interpretation, analysis, and time-lapse monitoring. In 2019, he toured the world presenting the Society of Exploration Geophysicists (SEG) Distinguished Lecture. In 2020, he was the recipient of the SEG Reginald Fessenden Award for his contributions to seismic data acquisition with compressive sensing. Since his arrival at Georgia Tech in 2017, he expanded his research program to include ML for Bayesian wave–equation-based inference using techniques from simulation-based inference. More recently, he started a research program on seismic monitoring of Geological Carbon Storage, which includes the development of an uncertainty-aware digital twin.

[1] Fortunato S. Community detection in graphs. Physics Reports. 2010;486(3–5):75–174.
[2] Fortunato S. A first course in network science. Cambridge University Press; 2020.

Bethany Lusch is a Computer Scientist in the AI/ML group at the Argonne Leadership Computing Facility at Argonne National Laboratory. Her research expertise includes developing methods and tools to integrate AI with science, especially for dynamical systems and partial differential equation-based simulations. Her recent work includes developing ML emulators to replace expensive parts of simulations, such as computational fluid dynamics simulations of engines and climate simulations. She is also working on methods that incorporate domain knowledge in ML, representation learning, and using ML to analyze supercomputer logs. She holds a Ph.D. and M.S. in applied mathematics from the University of Washington and a B.S. in mathematics from the University of Notre Dame.

Ryan McGranaghan is a Data Scientist and Research Scientist at the NASA Jet Propulsion Laboratory (JPL), where he works with the Machine Learning and Instrument Autonomy group to apply data science techniques robustly and responsibly to the earth and space sciences, to cultivate cross-NASA Center collaborations, and to explore more cohesive and plural scientific communities. He is also a core team member for the NASA Transformation to Open Science initiative, improving the accessibility, inclusivity, and reproducibility of science. His career, as in his life, centers around creating and cultivating transdisciplinary and transcommunity connections in pursuit of scientific discovery and flourishing.

Raphaël Pestourie earned his Ph.D. in Applied Mathematics and an A.M. in Statistics from Harvard University in 2020. Before attending Georgia Tech, he was a postdoctoral associate at MIT Mathematics, where he worked closely with the MIT-IBM Watson AI Lab. Raphaël's research focuses on scientific ML at the intersection of applied mathematics and ML and inverse design via scientific ML and large-scale electromagnetic design.

David Roy is a Professor in the Department of Geography, Environment and Spatial Sciences and the Director of the Center for Global Change and Earth Observations. He has a Ph.D. from the Department of Geography, Cambridge University, United Kingdom (1994), an M.Sc. degree in Remote Sensing and Image Processing from the Department of Meteorology, University of Edinburgh (1988), and a B.Sc. degree in Geophysics from the Department of Environmental Science, University of Lancaster (1987). In Europe, he held postdoctoral research fellowships at the Natural Environment Research Council Unit for Thematic Information Systems, University of Reading, United Kingdom, and at the Space Applications Institute, Joint Research Centre of the European Commission, Ispra, Italy. He moved to the United States in 1998 to join the Department of Geographical Sciences, University of Maryland (where he remains an Adjunct Professor), and led the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Data

Operational Product Evaluation group at NASA's Goddard Space Flight Center, for 8 years. He was then a Professor (and remains an Adjunct Professor) in the Geospatial Sciences Center of Excellence, South Dakota State University, for 13 years.

Steve Techtmann is an Associate Professor of Environmental Microbiology in the Department of Biological Sciences at Michigan Technological University. He is also the Associate Director of the Great Lakes Research Center at Michigan Tech, leading the Environmental Biotechnology initiatives. He completed his Ph.D. at the University of Maryland, Baltimore, and the Institute for Marine and Environmental Technology. Steve previously worked as a postdoctoral fellow at the Uniformed Services University as well as at the University of Tennessee, Knoxville, in the Center for Environmental Biotechnology. His research focuses on leveraging fundamental understanding of microorganisms and microbial communities to develop novel technologies for production of high-value products from waste streams as well as improve environmental monitoring.

Lingfei Wu is an Assistant Professor of Information Science at the University of Pittsburgh. His research uses big data, complexity science, and AI to explore Team Science and Innovation. Although prior studies highlight the benefits of teams, his work uniquely addresses their costs; specifically, how large teams can overshadow individual creativity and recognition. He also investigates organizational and policy changes to resolve conflicts between individual and team success in innovation. Lingfei's research is widely recognized, with publications in leading journals such as Nature and PNAS. His work has been featured in major outlets, including *The New York Times*, *Harvard Business Review*, *Forbes*, *The Atlantic*, and *Scientific American*. As a thought leader in research evaluation, Lingfei has consulted for organizations such as the National Institutes of Health, Novo Nordisk Foundation, and John Templeton Foundation. His research and teaching excellence have earned accolades such as the NSF CAREER Award, Richard King Mellon Award, and Oxford Martin Fellowship.

Appendix B. Additional Resources

Resource	Location	Category
C2OA2SE Website	https://www.mtu.edu/mtri/research/project-areas /machine-learning/c2oa2se-workshop/	Conference website (Dr. Kitchen)
JPL SUDS Workshop and Conference Pages	https://sudsconf.com/	Conference website (Dr. McGranaghan)
JPL SUDS Report	https://www.jpl.nasa.gov/go/suds/suds-report/	Report (Dr. McGranaghan)
Living Library of AI for Science	https://www.zotero.org/groups/5276219/science_ understanding_from_data_science_suds/library	Reference collection (Dr. McGranaghan)
SUDS Slack Space	https://join.slack.com/t/suds-community/shared_invite/ zt-2tpfe276f-njLHz~kDL8fwkwrldxfxBA	Slack channel (Dr. McGranaghan)
Call for Papers for Journal of Geophysical Research Machine Learning and Computation Special Issue	https://agupubs.onlinelibrary.wiley.com/hub/journal/ 29935210/homepage/call-for-papers/si-2024-000868	Journal call for papers (Dr. McGranaghan)
Seismic Laboratory for Imaging and Modeling	https://slim.gatech.edu/	Laboratory website (Dr. Herrmann)
Digital Shadow paper by Dr. Herrmann et al.	https://slim.gatech.edu/content/uncertainty-aware-digital-shadow-underground-multimodal-co2-storage-monitoring	Paper (Dr. Herrmann)
Dr. Lingfei Wu's personal website	https://lingfeiwu.github.io/	Personal website (Dr. Wu)

List of Symbols, Abbreviations, and Acronyms

AI	artificial intelligence
ARL	Army Research Laboratory
C2OA2SE	Cross-Disciplinary Challenges and Opportunities in AI Applications to Science and Engineering
CO_2	carbon dioxide
DEVCOM	U.S. Army Combat Capabilities Development Command
GPU	graphics processing unit
JPL	Jet Propulsion Laboratory
KOT	Koopman operator theory
LLM	large language model
ML	machine learning
MTRI	Michigan Tech Research Institute
NASA	National Aeronautics and Space Administration
NLP	Natural Language Processing
NSF	National Science Foundation
PDE	partial differential equation
PI	principal investigator
PNAS	Proceedings of the National Academy of Sciences
RL	reinforcement learning
SEG	Society of Exploration Geophysicists
STEM	Science, Technology, Engineering, and Mathematics
SUDS	Science Understanding from Data Science
TDA	topological data analysis

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLB CI TECH LIB