

2024 NSF HDR Ecosystem Conference Harvesting the Data Revolution



Report of Contributions

Contribution ID: 1

Type: **not specified**

Registration

Monday 9 September 2024 14:00 (4 hours)

The registration desk is in the Illinois Conference Center atrium, right in front of you when you enter through the main entrance via the parking lot.

Contribution ID: 2

Type: **not specified**

Welcome Reception

Monday 9 September 2024 18:00 (2 hours)

Contribution ID: 3

Type: **not specified**

Welcome and Introduction

Tuesday 10 September 2024 09:00 (15 minutes)

Presenter: NEUBAUER, Mark (Univ. Illinois at Urbana Champaign (US))

Contribution ID: 4

Type: **not specified**

Welcome from the Dean of the College of Engineering

Tuesday 10 September 2024 09:15 (15 minutes)

Presenter: BASHIR, Rashid (University of Illinois at Urbana-Champaign)

Contribution ID: 5

Type: **not specified**

NSF Welcome and HDR Overview

Tuesday 10 September 2024 09:30 (30 minutes)

Presenter: WALTON, Amy (National Science Foundation)

Contribution ID: 6

Type: **not specified**

Keynote talk: Knowledge-Guided Machine Learning: A New Framework for Accelerating Scientific Discovery and Addressing Global Environmental Challenges

Tuesday 10 September 2024 14:00 (1 hour)

Climate change, loss of bio-diversity, food/water/energy security for the growing population of the world are some of the greatest environmental challenges that are facing the humanity. These challenges have been traditionally studied by science and engineering communities via process-guided models that are grounded in scientific theories. Motivated by phenomenal success of Machine Learning (ML) in advancing areas such as computer vision and language modeling, there is a growing excitement in the scientific communities to harness the power of machine learning to address these societal challenges. In particular, massive amount of data about Earth and its environment is now continuously being generated by a large number of Earth observing satellites, in-situ sensors as well as physics-based models. These information-rich datasets in conjunction with recent ML advances offer huge potential for understanding how the Earth's climate and ecosystem have been changing, how they are being impacted by human actions, and for devising policies to manage them in a sustainable fashion. However, capturing this potential is contingent on a paradigm shift in data-intensive scientific discovery since the "black box" ML models often fail to generalize to scenarios not seen in the data used for training and produce results that are not consistent with scientific understanding of the phenomena.

This talk presents an overview of a new generation of machine learning algorithms, where scientific knowledge is deeply integrated in the design and training of machine learning models to accelerate scientific discovery. These knowledge-guided machine learning (KGML) techniques are fundamentally more powerful than standard machine learning approaches, and are particularly relevant for scientific and engineering problems that are traditionally addressed via process-guided (also called mechanistic or first principle-based) models, but whose solutions are hampered by incomplete or inaccurate knowledge of physics or underlying processes. While this talk will illustrate the potential of the KGML paradigm in the context of environmental problems (e.g., Ecology, Hydrology, Agronomy, climate science), the paradigm has the potential to greatly advance the pace of discovery in any discipline where mechanistic models are used.

Presenter: KUMAR, Vipin (University of Minnesota)

Contribution ID: 7

Type: **not specified**

A3D3

Tuesday 10 September 2024 10:30 (15 minutes)

Presenter: HSU, Shih-Chieh (University of Washington Seattle (US))

Session Classification: HDR Institute Overview, Activities and Accomplishments

Contribution ID: 8

Type: **not specified**

ID4

Tuesday 10 September 2024 10:45 (15 minutes)

Presenter: TOBERER, Eric (Colorado School of Mines)

Session Classification: HDR Institute Overview, Activities and Accomplishments

Contribution ID: 9

Type: **not specified**

Imageomics

Tuesday 10 September 2024 11:00 (15 minutes)

Presenter: BERGER-WOLF, Tanya (The Ohio State University)

Session Classification: HDR Institute Overview, Activities and Accomplishments

Contribution ID: **10**

Type: **not specified**

iHARP

Tuesday 10 September 2024 11:15 (15 minutes)

Presenter: JANEJA, Vandana (UMBC)

Session Classification: HDR Institute Overview, Activities and Accomplishments

Contribution ID: **11**

Type: **not specified**

I-GUIDE

Tuesday 10 September 2024 11:30 (15 minutes)

Presenter: PADMANABHAN, Anand (University of Illinois at Urbana-Champaign)

Session Classification: HDR Institute Overview, Activities and Accomplishments

Contribution ID: 12

Type: **not specified**

Public Lecture: Making AI Safe, Effective, and Trustworthy

Tuesday 10 September 2024 19:30 (1h 15m)

The data revolution has been replaced by the AI revolution. It seems that every day, we hear about a new area of human endeavor that has been conquered by AI. There are AI lawyers, AI doctors, AI artists, poets, and mathematicians. AI will predict what jobs we get, what medicines we should be treated with, and even how we should be educated.

If that's true, then AI policy should consist of getting the government out of the way and letting innovation bloom. Indeed, that is what some advocate. But in fact, sound AI policy rests on the same principles that sound science rests on: openness, transparency, and evidence. In order to ensure that all of us can benefit from innovations in AI, the best ideas in AI policy emphasize openness in research, transparency and accountability in the claims made by those seeking to deploy it, and above all a requirement that claims – bold claims at that – of efficacy are backed up by evidence that can be independently evaluated.

In this talk, I'll talk about how AI policy is trying to harness the AI revolution so that all of us, including those who have been traditionally left behind by tech innovation, can lead lives enriched rather than controlled by technology.

Presenter: VENKATASUBRAMANIAN, Suresh

Contribution ID: 13

Type: **not specified**

Breakout session summary

Thursday 12 September 2024 09:00 (1 hour)

Contribution ID: 14

Type: **not specified**

White Paper Planning

Thursday 12 September 2024 10:30 (30 minutes)

Presenters: NEUBAUER, Mark (Univ. Illinois at Urbana Champaign (US)); NEUBAUER, Mark Stephen (Univ. Illinois at Urbana-Champaign)

Contribution ID: 15

Type: **not specified**

Closeout & Advertisement of 4th HDR Ecosystem Conference

Thursday 12 September 2024 11:20 (20 minutes)

Presenters: NEUBAUER, Mark (Univ. Illinois at Urbana Champaign (US)); BOGHRAT, Diane (The Ohio State University)

Contribution ID: **16**

Type: **not specified**

Breakout #4

Contribution ID: 17

Type: **not specified**

Machine Learning Challenges Overview

Wednesday 11 September 2024 11:45 (15 minutes)

Presenter: HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US))

Session Classification: Machine Learning Challenges

Contribution ID: 18

Type: **not specified**

Detecting Novel Astrophysical Phenomena with Gravitational Waves (A3D3)

Wednesday 11 September 2024 12:00 (10 minutes)

Presenter: HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US))

Session Classification: Machine Learning Challenges

Contribution ID: 19

Type: **not specified**

Detecting Anomalous Climate Phenomena (iHARP)

Wednesday 11 September 2024 12:10 (10 minutes)

Presenter: GHOSH, Subhankar (University of Minnesota)

Session Classification: Machine Learning Challenges

Contribution ID: 20

Type: **not specified**

Anomaly Detection: Hybrid Butterflies (Imageomics)

Wednesday 11 September 2024 12:20 (10 minutes)

Presenter: CAMPOLONGO, Elizabeth (The Ohio State University)

Session Classification: Machine Learning Challenges

Contribution ID: 21

Type: **not specified**

Frontier AI for Science Security and Technology (FASST) initiative

Wednesday 11 September 2024 09:45 (30 minutes)

Presenter: CAPPELLO, Franck (Argonne National Laboratory)

Session Classification: AI & Data Infrastructure

Contribution ID: 22

Type: **not specified**

NSF OAC and the National Artificial Intelligence Research Resource (NAIRR) Pilot

Wednesday 11 September 2024 09:00 (45 minutes)

Presenter: ANTYPAS, Katerina (National Science Foundation)

Session Classification: AI & Data Infrastructure

Contribution ID: 23

Type: **not specified**

Knowledge Transfer in AI Workload Acceleration: from Academia to Industry

Wednesday 11 September 2024 10:45 (15 minutes)

At UIUC, there is strong ongoing collaboration between academia and industry, reflected in large industry-sponsored centers and institutes such as the IBM-Illinois Discovery Accelerator Institute (IIDAI) and the Center for Networked Intelligent Components and Environments (C-NICE), as well as in smaller and medium-sized partnerships like the AMD Center of Excellence and the Amazon-Illinois Center on AI for Interactive Conversational Experiences (AICE). The research activities in these centers are closely aligned with the strategic goals of industry partners, driving innovation, performance improvements, and market competitiveness across various high-tech sectors. Through close collaboration between UIUC students, faculty, and industry professionals, we aim to publish technical papers in top conferences and journals, while also transferring valuable knowledge back to industry. Additionally, many of these projects are open-source, with code freely available to benefit both the research community and industry. In this talk, Prof. Chen will highlight these topics, with a focus on AI workload acceleration.

Presenter: CHEN, Deming (University of Illinois at Urbana-Champaign)

Contribution ID: 24

Type: **not specified**

Communicating Science: Effective Delivery To Any Audience

Wednesday 11 September 2024 14:00 (30 minutes)

Abstract:

Effective science communication is crucial for bridging the gap between complex scientific concepts and diverse audiences. In this talk, I will explore strategies to convey scientific ideas clearly and engagingly, ensuring accessibility for individuals with varying levels of background knowledge. We will discuss the importance of storytelling in science communication, as well as techniques for simplifying technical jargon without compromising the integrity of the information. This talk aims to empower scientists, educators, and communicators to effectively share their work with broader, more diverse audiences, fostering a greater public understanding and appreciation of science.

Bio:

Sara Ayman Metwalli has a Ph.D. in Quantum Computing at Keio University, Japan, where she focused on developing and optimizing quantum algorithms and debugging tools for Noisy Intermediate-Scale Quantum (NISQ) devices. Sara has made significant contributions to quantum information science, particularly in the areas of quantum error correction and fault-tolerance. Her work has been published in leading journals and presented at international conferences, highlighting her role as an emerging leader in the quantum computing field.

In addition to her research, Sara has a strong commitment to STEM education and outreach. She has extensive experience as an educator, teaching programming and quantum computing to a wide range of students, from K-12 to university graduates. Sara is passionate about increasing diversity in STEM and has actively worked to create inclusive educational environments that empower underrepresented groups to pursue careers in science and technology.”

Presenter: METWALLI, Sara (Argonne National Laboratory)

Contribution ID: 25

Type: **not specified**

NSF Panel

Tuesday 10 September 2024 11:45 (45 minutes)

Moderator: Tanya Berger-Wolf (The Ohio State University)

A panel of the NSF program directors bringing the NSF perspective of the history and the future of the data revolution, particularly in the context of AI revolution.

Panelists primarily assembled from Cognizant Program Officers of NSF 21-519 HDR Institute and NSF24-560 DSC:

- Amy Walton** Deputy Director for the Office of Advanced Cyberinfrastructure (OAC) at the National Science Foundation
- Sylvia Spengler**, Program Director, Division of Information and Intelligent System (IIS) within the Computer and Information Science and Engineering (CISE) Directorate
- Chaitanya K. Baru**, Senior Advisor, Directorate for Technology, Innovation and Partnerships (TIP)
- Raleigh Martin** Program Director, Directorate for Geosciences (GEO) Division of Earth Sciences (EAR) Integrated Activities Section
- Cheryl L. Eavey** Directorate for Social, Behavioral and Economic Sciences (SBE) Division of Social and Economic Sciences (SES) Methodology, Measurement, and Statistics (MMS)

Contribution ID: 26

Type: **not specified**

Dinner on your own

Contribution ID: 27

Type: **not specified**

Promoting Educational and Outreach Activities throughout the HDR Institutes via a Centralized Repository

Wednesday 11 September 2024 14:30 (10 minutes)

During the ideation expo at the previous HDR Ecosystem Conference in Colorado (Oct 2023), the creation of an HDR-wide educational and outreach materials repository was proposed, with a committee of cross-institutional members formed to build this repository. The long-term vision for the repository is to promote and facilitate data fluency in domain specific contexts for communities ranging from K-12th grade classrooms to the general public. In this talk, I will describe our efforts to design and implement a centralized repository to host educational and outreach materials collected throughout the HDR institutes. I will also report on a recent workshop (with cross-institutional participation) on the development of data fluency learning modules targeting 6-12th grade STEM classrooms, which we have included in the repository. Finally, I will describe our plans moving forward with the eventual goal of public promotion and release.

Presenter: PAK, Alex (Colorado School of Mines)

Contribution ID: 28

Type: **not specified**

Pre-tenure faculty & postdoc mentoring lunch

Wednesday 11 September 2024 12:30 (1h 30m)

Contribution ID: 29

Type: **not specified**

Conference Summary

Thursday 12 September 2024 11:00 (20 minutes)

Presenter: HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US))

Contribution ID: **30**

Type: **not specified**

Imageomics: FAIR ML Products for Biological Knowledge Discovery

Tuesday 10 September 2024 17:35 (1 minute)

Presenter: CAMPOLONGO, Elizabeth

Session Classification: Poster Session

Contribution ID: 31

Type: **not specified**

Incorporating phenotypic similarity into trait description embeddings

Tuesday 10 September 2024 17:36 (1 minute)

Presenter: KAR, Soumyashree

Session Classification: Poster Session

Contribution ID: 32

Type: **not specified**

Latent Space Phenotyping for Measuring Complex Evolutionary Traits

Tuesday 10 September 2024 17:37 (1 minute)

Presenter: CHARPENTIER, Caleb

Session Classification: Poster Session

Contribution ID: 33

Type: **not specified**

Hierarchical Conditioning of Diffusion Models Using Tree-of-Life for Studying Species Evolution

Tuesday 10 September 2024 17:38 (1 minute)

Presenter: KHURANA, Mridul

Session Classification: Poster Session

Contribution ID: 34

Type: **not specified**

What Do You See in Common? Learning Hierarchical Prototypes over Tree-of-Life to Discover Evolutionary Traits

Tuesday 10 September 2024 17:39 (1 minute)

Presenter: BABU MANOGARAN, Harish

Session Classification: Poster Session

Contribution ID: 35

Type: **not specified**

BioCLIP: A Vision Foundation Model for the Tree of Life

Tuesday 10 September 2024 17:40 (1 minute)

Presenter: STEVENS, Sam

Session Classification: Poster Session

Contribution ID: 36

Type: **not specified**

Education and Outreach in Imageomics: Engaging Communities to Advance Science

Tuesday 10 September 2024 17:41 (1 minute)

Presenter: BOGHRAT, Diane

Session Classification: Poster Session

Contribution ID: 37

Type: **not specified**

Using Deep Learning to Quantify Phenotypic Similarities in Mimic Butterfly Species using Human, Bird, and Butterfly Acuties

Tuesday 10 September 2024 17:42 (1 minute)

Presenter: RAMIREZ, Michelle

Session Classification: Poster Session

Contribution ID: **38**

Type: **not specified**

Dynamic Network Classification

Tuesday 10 September 2024 17:43 (1 minute)

Presenter: BANERJI, Namrata

Session Classification: Poster Session

Contribution ID: 39

Type: **not specified**

Tulane Center for Community-Engaged Artificial Intelligence

Tuesday 10 September 2024 17:44 (1 minute)

Presenter: CULOTTA, Aron

Session Classification: Poster Session

Contribution ID: 40

Type: **not specified**

Understanding of impact of training size on animal re-identification

Tuesday 10 September 2024 17:45 (1 minute)

Presenter: NEPOVINNYKH, Ekaterina

Session Classification: Poster Session

Contribution ID: 41

Type: **not specified**

Practical Leadership for Team Science: Experiences from the Imageomics Institute

Tuesday 10 September 2024 17:46 (1 minute)

Presenter: BOGHRAT, Diane

Session Classification: Poster Session

Contribution ID: 42

Type: **not specified**

VLM4Bio: A Benchmark Dataset to Evaluate Pretrained Vision-Language Models for Trait Discovery from Biological Images

Tuesday 10 September 2024 17:47 (1 minute)

Presenter: KARPATNE, Anuj

Session Classification: Poster Session

Contribution ID: 43

Type: **not specified**

What Do You See in Common? Learning Hierarchical Prototypes over Tree-of-Life to Discover Evolutionary Traits

Tuesday 10 September 2024 17:48 (1 minute)

Presenter: KARPATNE, Anuj

Session Classification: Poster Session

Contribution ID: 44

Type: **not specified**

National Data Mine Network

Tuesday 10 September 2024 17:49 (1 minute)

Presenter: DANIEL WARD, Mark

Session Classification: Poster Session

Contribution ID: 45

Type: **not specified**

Fish-Vista: A Multi-Purpose Dataset for Understanding & Identification of Traits from Images

Tuesday 10 September 2024 17:50 (1 minute)

Presenter: SAJEED MEHRAB, Kazi

Session Classification: Poster Session

Contribution ID: 46

Type: **not specified**

Low-cost Efficient Wireless Intelligent Sensors (LEWIS) for Engineering Education: Native American Knowledge for Data Science Education

Tuesday 10 September 2024 17:51 (1 minute)

Presenter: MOREU, Fernando

Session Classification: Poster Session

Contribution ID: 47

Type: **not specified**

Facilitating Knowledge Sharing and Discovery: Search Functionality and API Design for the I-GUIDE

Tuesday 10 September 2024 17:52 (1 minute)

Presenter: KANG, Yunfan

Session Classification: Poster Session

Contribution ID: 48

Type: **not specified**

CLV: A Novel Framework for Enhanced Anomaly Detection and Attribution in Multivariate Time Series Data

Tuesday 10 September 2024 17:53 (1 minute)

Presenter: ALE, Tolulope

Session Classification: Poster Session

Contribution ID: 49

Type: **not specified**

Battling Misinformation through Interdisciplinary Collaboration

Tuesday 10 September 2024 17:54 (1 minute)

Presenter: KHANJANI, Zahra

Session Classification: Poster Session

Contribution ID: 50

Type: **not specified**

CMAD: Advancing Understanding of Anomalous Melt Events over the Antarctic Sea Ice

Tuesday 10 September 2024 17:55 (1 minute)

Presenter: KUMAR DEVNATH, Maloy

Session Classification: Poster Session

Contribution ID: 51

Type: **not specified**

HDR DSC: The Metropolitan Chicago Data-science Corps (MCDC)

Tuesday 10 September 2024 17:56 (1 minute)

Presenter: SHI, Lizhen

Session Classification: Poster Session

Contribution ID: 52

Type: **not specified**

Neural Network Efficiency Evaluation on the AMD Versal AI Engine

Tuesday 10 September 2024 17:57 (1 minute)

Presenter: SHEN, Yilin

Session Classification: Poster Session

Contribution ID: 53

Type: **not specified**

BaboonLand Dataset: Tracking Primates in the Wild and Automating Behaviour Recognition from Drone Videos

Tuesday 10 September 2024 17:58 (1 minute)

Presenter: KHOLIIVCHENKO, Maksim

Session Classification: Poster Session

Contribution ID: 54

Type: **not specified**

Assessing Annotation AccAssessing Accuracy in Ice Sheets Using Quantitative Metrics

Tuesday 10 September 2024 17:59 (1 minute)

Presenter: TAMA, Bayu

Session Classification: Poster Session

Contribution ID: 55

Type: **not specified**

Genotype to Phenotype Mapping via Deep Learning

Tuesday 10 September 2024 18:00 (1 minute)

Presenter: CARLYN, David

Session Classification: Poster Session

Contribution ID: 56

Type: **not specified**

Variance Analysis of Brightness Temperature using High-resolution DYAMOND simulations and CRTM in Digital Twin Systems

Tuesday 10 September 2024 18:01 (1 minute)

Presenter: KULKARNI, Chhaya

Session Classification: Poster Session

Contribution ID: 57

Type: **not specified**

Predicting Sea ice extent over Antarctica using Patch CNN

Tuesday 10 September 2024 18:02 (1 minute)

Presenter: VIKAS AMARANENI, Sai

Session Classification: Poster Session

Contribution ID: 58

Type: **not specified**

HDR DSC: Collaborative Research: Transforming Data Science Education through a Portable and Sustainable Anthropocentric Data Analytics for Community Enrichment (ADACE) Program

Tuesday 10 September 2024 18:03 (1 minute)

Presenter: LIANG, Yu

Session Classification: Poster Session

Contribution ID: 59

Type: **not specified**

Physics-Informed Sea Ice Thickness Prediction

Tuesday 10 September 2024 18:04 (1 minute)

Presenter: SAMPATH, Akila

Session Classification: Poster Session

Contribution ID: **60**

Type: **not specified**

Probabilistic Prediction of Material Stability: Integrating Convex Hulls into Active Learning

Tuesday 10 September 2024 18:05 (1 minute)

Presenter: NOVICK, Andrew

Session Classification: Poster Session

Contribution ID: 61

Type: **not specified**

Cyberinfrastructure for Scientific Data Preservation and Image Similarity Search

Tuesday 10 September 2024 18:06 (1 minute)

Presenter: AGAR, Joshua

Session Classification: Poster Session