**[AI4S] 5th International Workshop on Artificial Intelligence and**

**Machine Learning for Scientific Applications**

**Held in conjunction with SC24**

**November 18, 2024**

It is our great pleasure to welcome you to AI4S’24, the 5th International Workshop on Artificial Intelligence and Machine Learning for Scientific Applications. Artificial intelligence (AI) and machine learning (ML) are transformative technologies that have revolutionized the landscape of algorithms, implementations, and applications across various fields. The concept of “AI for Science” encompasses the creation of future methodologies and scientific breakthroughs through the integration of computational learning and machine intelligence. This includes the development and application of AI techniques, such as machine learning, deep learning, statistical methods, data analytics, automated control, and related areas. We have witnessed numerous success stories where AI methods have been applied to predict extreme weather events, identify exoplanets within vast amounts of sky data, accelerate fluid simulation solvers, design advanced materials, fast-track drug discovery, explore the mysteries of the universe, and drive many other groundbreaking scientific discoveries. Despite these advancements, several challenges remain in enhancing the usability of AI/ML in scientific applications, particularly when leveraging high-performance computing (HPC) systems. Key questions include: How can we systematically and automatically apply AI/ML to scientific applications? How do we incorporate domain-specific knowledge, such as conservation laws, invariants, causality, and symmetries, into AI/ML models? How can we ensure these models are interpretable, robust, and suitable for HPC environments? How do we make AI/ML approaches more accessible to the HPC community? And how can we effectively harness extreme-scale HPC systems and novel AI accelerators for AI/ML? Addressing these challenges will bridge the gap between AI/ML and scientific applications, enabling more widespread adoption and unlocking the full potential of AI/ML in HPC environments.

The main challenge of AI4S is to bring together researchers from academia, industry, and government authorities to collaborate on various aspects of AI/ML in scientific applications. One of the primary goals is to foster meaningful interaction among these stakeholders, enabling them to share recent advances in applying AI/ML to a wide range of scientific domains. This workshop also aims to introduce new scientific challenges to the broader community and encourage the development of tools and infrastructures that support the integration of AI/ML into scientific workflows. By facilitating these collaborations, AI4S seeks to drive innovation and enhance the effectiveness of AI/ML in advancing scientific discovery.

In this fifth edition of AI4S, after an intensive review process, the workshop features a program that includes 12 papers. Additionally, the program boasts invited talks from two distinguished speakers: Torsten Hoefler and Satoshi Matsuoka. Torsten Hoefler is a full professor at ETH Zurich, where he leads the Scalable Parallel Computing Laboratory (SPCL). He also serves as the Chief Architect for Machine Learning at the Swiss National Supercomputing Center and is a long-term consultant for Microsoft, specializing in large-scale AI and networking. Satoshi Matsuoka has been the director of the RIKEN Center for Computational Science (R-CCS) since 2018. He played a key role in developing Fugaku, the world’s fastest supercomputer in 2020 and 2021, dominating all four major supercomputer rankings (Top500, HPCG, HPL-AI, Graph500). His work extends to cutting-edge research in high-performance computing (HPC), including investigating Post-Moore era computing and the future Fugaku NEXT supercomputer.

We would like to extend our sincere thanks to everyone who worked hard to make this workshop possible. Our gratitude goes to the authors and invited speakers for providing the content of our program and sharing their valuable contributions at the conference. We also want to thank the Program Committee members for their diligent reviews, insightful feedback, and assistance in shaping the workshop’s program. We hope this workshop fosters meaningful collaboration and idea-sharing among researchers and practitioners from around the world, contributing to the growth of this emerging community in such an important field of research.

Please visit https://ai4s.github.io/ to learn more about this workshop.

**Workshop Organizers**

* Gokcen Kestor, Pacific Northwest National Laboratory, USA, [gokcen.kestor@pnnl.gov](mailto:gokcen.kestor@pnnl.gov)
* Dong Li, University of California, Merced, USA, [dli35@ucmerced.edu](mailto:dli35@ucmerced.edu)
* Murali Emani, Argonne National Laboratory, USA, [memani@anl.gov](mailto:memani@anl.gov)

**Workshop Committee**

* Bilge Acun, Meta
* Jeff Daily, AMD
* Aleksandr Drozd, RIKEN
* Akash Dutta, AMD
* Rio Yokota, Tokyo Institute of Technology
* Väinö Hatanpää, ANL
* Kevin J. Barker, Pacific Northwest National Laboratory
* Mehmet E Belviranli, Colorado School of Mines
* Wesley Brewer, Oak Ridge National Laboratory
* Wenqian Dong, Oregon State University
* Aleksandr Drozd, RIKEN
* Steven Farrell, Lawrence Berkeley National Lab
* Roberto Gioiosa, Pacific Northwest National Lab
* Jiajia Li, North Carolina State University
* Zhengchun Liu, AWS
* Jacques Pienaar, Google
* Ben Ren, William and Mary
* Kento Sato, RIKEN
* Bin Ren, William&Mary
* Dingwen Tao, Indiana University
* Brian C. Van Essen, Lawrence Livermore National Laboratory
* Venkatram Vishwanath, Argonne National Laboratory
* Feiyi Wang, Oak Ridge National Laboratory
* Bo Yuan, Rutgers University

**Workshop Agenda**

**Invited Talk 1: Torsten Hoefler (ETH Zurich)**

**Paper session – 1**

* Machine Learning Aboard the ADAPT Gamma-Ray Telescope

Ye Htet, Marion Sudvarg, Andrew Butzel, Jeremy Buhler, Roger Chamberlain, James Buckley

* A Scalable Real-Time Data Assimilation Framework for Predicting Turbulent Atmosphere Dynamics

Feng Bao, Hristo G. Chipilski, Siming Liang, Siyan Liu, Dan Lu, Junqi Yin, Guannan Zhang

* ChatBLAS: The First AI-Generated and Portable BLAS Library

Pedro Valero-Lara, William Godoy, Keita Teranishi, Prasanna Balaprakash, Jeffrey Vetter

**Paper session – 2**

* EchoStateNetworks: A Non-Intrusive Approach to Anomaly Detection in Manufacturing

Kendric Hood

* MelissaDL x Breed: Towards Data-Efficient On-Line Supervised Training of Multi-Parametric Surrogates with Active Learning

Sofya Dymchenko, Abhishek Purandare, Bruno Raffin

* Fourier neural operators for spatiotemporal dynamics in two-dimensional turbulence

Mohammad Atif, Pulkit Dubey, Pratik P. Aghor, Vanessa Lopez-Marrero, Tao Zhang, Abdullah Sharfuddin, Kwangmin Yu, Fan Yang, Foluso Ladeinde, Yangang Liu, Meifeng Lin, Lingda Li

**Lunch Break**

**Invited Talk 2: Satoshi Matsuoka (RIKEN, Japan)**

**Paper session – 3**

* ChatVis: Automating Scientific Visualization with a Large Language Model

Tanwi Mallick, Orcun Yildiz, David Lenz, Tom Peterka

* A Comparative Survey: Reusing Small Pre-Trained Models for Efficient Large Model Training

 Dhroov Pandey, Jonah Ghebremichael, Zongqing Qi, Tong Shu

* Enhancing Electron Microscopy Image Classification Using Data Augmentation

Jordan Alan Welsman, Gunther H. Weber, Oluwamayowa O. Amusat, Anna Giannakou, Lavanya Ramakrishnan

**Paper session – 4**

* SciTrust: Evaluating the Trustworthiness of Large Language Models for Science

Emily Herron, Junqi Yin, Feiyi Wang

* AI Surrogate Model for Distributed Computing Workloads

David K. Park, Yihui Ren, Ozgur O. Kilic, Tatiana Korchuganova, Sairam Sri Vatsavai, Joseph Boudreau, Tasnuva Chowdhury, Shengyu Feng, Raees A. Khan,  Jaehyung Kim, Scott Klasky, Tadashi Maeno, Paul Nilsson,  Verena Ingrid Martinez Outschoorn, Norbert Podhorszki, Frédéric Suter,  Wei Yang,  Yiming Yang, Shinjae Yoo, Alexei Klimentov, Adolfy Hoisie

* AstroMLab 2: Benchmarking Specialised LLMs for Astronomy

Rui Pan, Tuan Dung Nguyen, Alberto Accomazzi, Tirthankar Ghosal,  Yuan-Sen Ting