

The 1st International Workshop on Data-driven AI (DATAI)

DATAI 2024

Abstract

The advent of artificial intelligence (AI), particularly through deep learning (DL) and large language models (LLMs), has marked a significant milestone in technological advancement, attributing to its unparalleled accuracy and generalization abilities. The rapid evolution of AI model structures to achieve superior performance underscores the dynamic progression and potential of AI technologies. However, the cornerstone of any AI's success lies not just in its algorithmic prowess but in the quality of data it is trained on. High-quality, accurate, consistent, and representative data sets are imperative for enhancing AI models' learning efficacy, thereby optimizing their generalization capabilities and reducing computational demands.

The 1st International Workshop on Data-driven AI (DATAI) is dedicated to exploring the symbiotic relationship between AI technologies and high-quality data. It aims to delve into the latest research breakthroughs and present innovative techniques and methodologies at the forefront of Data-driven AI. This workshop will provide a comprehensive platform to discuss the construction of high-quality data specifically tailored for AI technologies, focusing on large-scale models. It seeks to engage researchers, developers, and practitioners in rigorous discussions to explore sustained advancements, design innovations, and practical applications of data construction techniques that drive AI technologies forward.

Topics of Interest:

- Data discovery for ML
- Data cleaning & integration for ML
- Labeling quality and ML performance
- Data-efficient solutions for ML training
- LLM-based data cleaning & integration

By fostering a collaborative environment, DATAI aims to inspire a diverse audience of participants from the realms of AI and data quality management, facilitating an exchange of ideas that propels the field toward groundbreaking developments.

Important Dates

- Workshop Website Availability and Call for Papers: February 20, 2024
- Submission Deadline for Research Papers: April 1, 2024
- Notification of Authors: April 15, 2024
- Camera-ready Version of Accepted Papers: April 27, 2024

Organizing Team

General Chairs:

Hongzhi Wang, Professor and PhD supervisor at Harbin Institute of Technology.

Secretary General of ACM SIGMOD China and an outstanding member of CCF. Research focuses include big data management, analysis, and data quality. Nan Tang, Associate Professor at the University of Science and Technology, Guangzhou. Notable for his senior scientist role at Qatar Computing Research Institute and research fellow at University of Edinburgh.

PC Chairs:

Lei Cao, Assistant Professor at University of Arizona and MIT CSAIL. Research interests in data systems and science.

Chengliang Chai, Associate Professor in Computer Science and Technology at Beijing Institute of Technology. Focuses on leveraging data management techniques for AI.

Xiaoou Ding, Assistant Professor in Computer Science and Technology, Harbin Institute of Technology. Specializes in data cleaning, integration, and temporal data quality management.

Publication

All accepted papers are required to strictly follow the formatting guidelines as per the Camera Ready instructions outlined by VLDB. These guidelines are crucial for maintaining the integrity and uniformity of the workshop proceedings. Authors must ensure their submissions are prepared according to the instructions available at [VLDB Volume 17 Formatting Guidelines](#). Compliance with these formatting requirements is essential, as the template cannot enforce them. The final papers will be rigorously reviewed by the workshop PC members to ensure adherence to these standards before forwarding them to the VLDB workshop proceedings chairs for publication.

Program Committee

The workshop's program committee comprises a distinguished group of experts in data management, analytics, data science, and machine learning, tasked with reviewing submissions and ensuring high-quality research presentations. Notable members include:

- **Dong Deng** – Rutgers University
- **Raul Fernandez** – University of Chicago, USA
- **El Kindi Rezig** – University of Utah, USA
- **Ju Fan** – Remin University
- **Jia Zou** – ASU, USA
- **Jiannan Wang** – Simon Fraser University
- **Meihui Zhang** – Beijing Institute of Technology
- **Paolo Papotti** – Eurocom, France
- **Mourad Ouzzani** – QCRI, Qatar
- **Shaoxu Song** – Tsinghua University
- **Xu Chu** – Georgia Institute of Technology, USA
- **Ziawasch Abedjan** – Leibniz Universität Hannover, Germany
- **Oscar Moll** – MIT, USA
- **Peter Chen** – MIT, USA

Contact Information:

- Hongzhi Wang: wangzh@hit.edu.cn
- Nan Tang: nantang@hkust-gz.edu.cn