

Seventh International Workshop on Exploiting Artificial Intelligence Techniques for Data Management (aiDM)

Rajesh Bordawekar
IBM T.J. Watson Research Center
Yorktown Heights, NY, United States
bordaw@us.ibm.com

Oded Shmueli
Technion - Israel Institute of
Technology
Haifa, Israel
oshmu@cs.technion.ac.il

Yael Amsterdamer
Bar-Ilan University
Ramat Gan, Israel
first.last@biu.ac.il

Renata Borovica-Gajic
The University of Melbourne
Melbourne, Australia
renata.borovica@unimelb.edu.au

Donatella Firmani
Sapienza University of Rome
Rome, Italy
donatella.firmani@uniroma1.it

ABSTRACT

Recent advances in AI techniques, as well as enabling hardware and infrastructure, have led to the integration of AI across wide-ranging domains and tasks. In particular, AI has been used to handle various types of data (including numerical, textual and image data) and has been adopted in large-scale distributed systems. From a data management perspective, this calls for the harnessing of state-of-the-art AI solutions for data management tasks and systems. aiDM is a full-day workshop that offers a stage for innovative interdisciplinary research that studies the interaction between AI and data management and develops new AI technologies for data-related tasks. This year, aiDM'24 particularly focuses on the transparent exploitation of AI techniques (e.g., using Generative AI frameworks) for data management for enterprise class workloads.

CCS CONCEPTS

• **Information systems** → **Database management system engines**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Data management; Artificial intelligence; Machine learning

ACM Reference Format:

Rajesh Bordawekar, Oded Shmueli, Yael Amsterdamer, Renata Borovica-Gajic, and Donatella Firmani. 2024. Seventh International Workshop on Exploiting Artificial Intelligence Techniques for Data Management (aiDM). In *Companion of the 2024 International Conference on Management of Data (SIGMOD-Companion '24)*, June 9–15, 2024, Santiago, AA, Chile. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3626246.3655013>

1 INTRODUCTION

Recently, the Artificial Intelligence (AI) field has been experiencing a resurgence. AI broadly covers a wide swath of techniques, which include logic-based approaches, probabilistic graphical models, and machine learning/deep learning approaches. Advances in hardware

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMOD-Companion '24, June 9–15, 2024, Santiago, AA, Chile

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0422-2/24/06.

<https://doi.org/10.1145/3626246.3655013>

capabilities, such as Graphics Processing Units (GPUs), software components (e.g., accelerated libraries, programming frameworks), and systems infrastructures (e.g., GPU-enabled cloud providers) have led to a wide-spread adaptation of AI techniques to a variety of domains. Examples of such domains include image classification (e.g., [7]), autonomous driving (e.g., [4]), automatic speech recognition (ASR, e.g., [6]) and conversational systems (chatbots, e.g. [8]). AI techniques not only support multiple datatypes (e.g., free text, images, or speech), but are also available in various configurations, from personal devices to large-scale distributed systems.

In spite of the wide ranging applications of AI techniques, its interactions with the data management systems remains in infancy. At present, in the context of AI, a majority of database management systems (DBMS) are being used primarily as a repository for feeding input data and storing results. Recently, there has been some activity in using AI techniques in DBMSs, e.g., enabling natural language interfaces to relational databases (e.g., [1]) and applying machine learning techniques for query optimizations, performance tuning or learned index structures (e.g., [2, 3, 5]). Nonetheless, a lot more work is needed to harness the power of AI in data management.

The aiDM workshop brings together people from academia and industry to discuss various ways of integrating AI techniques with data management systems. The primary goal of the workshop is to explore opportunities for AI techniques for enhancing different components of data management systems, e.g., user interfaces, tooling, performance optimizations, new query types, and workloads. Special emphasis will be given to transparent exploitation of AI techniques using existing data management for enterprise class workloads. We aim to identify important areas of research and spur new efforts in this emerging field.

2 TOPICS OF INTEREST

The goal of the workshop is to take a holistic view of various AI technologies and investigate how they can be applied to different components of an end-to-end data management pipeline. Special emphasis will be given to how AI techniques could be used for enhancing user experience by reducing complexity in tools, providing newer insights, or providing better user interfaces. Topics of interest include, but are not restricted to:

- AI-enabled improvements to foundational DB algorithms: sorting, searching, consensus

- New AI-enabled business intelligence (BI) queries for relational databases
- Integration of Large Language Models with databases and supporting services (e.g., Generative AI)
- Natural language queries and conversational interfaces
- AI-enabled database programming (e.g., natural language queries, SQL co-pilots, etc.)
- Design and Implementation of Vector Databases for unstructured data
- Ethics, governance, and societal implications of AI-enabled databases
- Reasoning over knowledge bases
- Self-tuning Databases using reinforced learning
- Impact of model interpretability
- Supporting multi-modality (e.g., images or time-series data)
- Supporting semi-structured, streaming, and graph databases
- AI-enabled data integration and data imputation strategies
- Impact of AI on tooling, e.g., ETL or data cleaning
- Performance implications of AI-enabled queries
- AI-enabled databases for managing/supporting AI workloads
- AI strategies for data provenance, access control, anomaly detection and cyber security

As this area is new and fluid, we encourage submissions that describe on-going projects as well.

2.1 Related workshops

We view SIGMOD DEEM as a workshop to examine databases for Machine Learning (ML) issues, i.e., DEEM is focused on infrastructure for ML workloads. ICDE DBML also focuses on the integration of DBs and ML. In contrast, our focus is on AI for data management – not just for traditional ML. AI is a related but distinct topic, and given the current focus on Generative AI, these differences have become even more prominent. We want to explore different AI techniques (e.g., learning, probabilistic modeling, NLP, etc.) that can be used in various database operations such as queries, utilities, data exploration, visualization, etc. Specifically, we are interested in exploring how AI can enhance the capabilities/ experiences of data management systems. At its inception in SIGMOD'18, the aiDM workshop was the first workshop dedicated to exploration of AI techniques for database workloads. Since then, it has emerged as the premier workshop on this topic, with consistent interest from the broader research community (since inception, aiDM papers have more than 10,000 downloads from the ACM digital library). DATAI (Data Driven AI) workshop, being co-hosted with the VLDB conference, deals with discovering and cleaning data to be used by AI models, and hence serves a different purpose than our workshop.

3 PROGRAM AND ORGANIZATION

aiDM 2024 is a full-day workshop with three keynote presentations (academic, industrial and practitioner) and/or a panel discussion and presentations of accepted papers, both individually and at a common poster session. Full details on the program and the accepted papers will be available at the workshop website (aidm-conf.org).

3.1 Reviewing Process

Each submitted paper will be reviewed by at least three reviewers, using the same conflict rules as SIGMOD 2024 (e.g., double-anonymous reviewing and conflict declarations). The final selection

is done by the workshop chairs, based on collective review scores. We will accept both long (8 pages) and short (4 pages) submissions. The workshop organizers are not allowed to submit at the workshop and the program committee members are strongly discouraged.

3.2 Program Committee

We would like to thank the program committee members for their efforts and valuable reviews and comments.

- Madelon Hulsebos, UC Berkeley
- Sonia-Florina Horchidan, KTH Royal Institute of Technology
- Yuliang Li, Reality Lab Research, Meta
- Yasuko Matsubara, Osaka University
- Apoorva Nitsure, IBM Almaden Research Center
- Amit Somech, Bar-Ilan University
- Matthias Urban, TU Darmstadt
- Jun Wan, Databricks
- Wenlu Wang, Texas A&M University

3.3 Organizers

The aiDM'24 workshop planning will be guided by a steering committee consisting of

- Rajesh Bordawekar, IBM T. J. Watson Research Center
- Oded Shmueli, Technion

The workshop program chairs are

- Yael Amerdamer, Bar-Ilan University
- Renata Borovica-Gajic, The University of Melbourne
- Donatella Firmani, Sapienza University of Rome

ACKNOWLEDGMENTS

Some authors are financially supported by the Israel Science Foundation (grant no. 2015/21) and a grant from the Israel Ministry of Science and Technology. Some authors are partially supported by EU Horizon 2020 project INTEND (grant no. 101135576), and some are supported by the Australian Research Council DECRA (grant no. DE230100366), and L'Oreal for Women in Science'23 Fellowship.

REFERENCES

- [1] Katrin Affolter, Kurt Stockinger, and Abraham Bernstein. 2019. A comparative survey of recent natural language interfaces for databases. *VLDB J.* 28, 5 (2019), 793–819.
- [2] Hai Lan, Zhifeng Bao, J. Shane Culpepper, and Renata Borovica-Gajic. 2023. Updatable Learned Indexes Meet Disk-Resident DBMS - From Evaluations to Design Choices. *Proceedings of the ACM on Management of Data (SIGMOD)* 1, 2 (2023), 139:1–139:22.
- [3] Ryan Marcus, Parimarjan Negi, Hongzi Mao, Nesime Tatbul, Mohammad Alizadeh, and Tim Kraska. 2021. Bao: Making Learned Query Optimization Practical. In *SIGMOD*. 1275–1288.
- [4] Khan Muhammad, Amin Ullah, Jaime Lloret, Javier Del Ser, and Victor Hugo C. de Albuquerque. 2021. Deep Learning for Safe Autonomous Driving: Current Challenges and Future Directions. *IEEE Trans. Intell. Transp. Syst.* 22, 7 (2021), 4316–4336.
- [5] R. Malinga Perera, Bastian Oetomo, Benjamin I. P. Rubinstein, and Renata Borovica-Gajic. 2023. No DBA? No Regret! Multi-Armed Bandits for Index Tuning of Analytical and HTAP Workloads With Provable Guarantees. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* 35, 12 (2023), 12855–12872.
- [6] Yao Qin, Nicholas Carlini, Garrison W. Cottrell, Ian J. Goodfellow, and Colin Raffel. 2019. Imperceptible, Robust, and Targeted Adversarial Examples for Automatic Speech Recognition. In *ICML*. 5231–5240.
- [7] Waseem Rawat and Zenghui Wang. 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.* 29, 9 (2017), 2352–2449.
- [8] Rui Yan. 2018. "Chitty-Chitty-Chat Bot": Deep Learning for Conversational AI. In *IJCAI*. 5520–5526.