

## Preface

Database and Artificial Intelligence (AI) can benefit from each other. On the one hand, AI can make database more intelligent (AI4DB) by exploiting learning-based techniques. On the other hand, database techniques can optimize AI models (DB4AI), such as reducing the complexity of using AI models and accelerating the deployment of AI algorithms. In this special section, we discuss 1) how to exploit AI or machine learning techniques for index design, performance tuning, query processing in database systems, and 2) how to utilize database and data management techniques to make AI models more reusable and more tolerant to dirty data.

To promote the recent work on AI4DB and DB4AI, we organize this special section at Journal of Computer Science and Technology (JCST). We received 20 papers from all over the world. First, the guest editors preformed quick reviews and immediately rejected those of not sufficiently high quality submissions. Then, each remaining submission was reviewed by at least three invited international reviewers. All the papers were carried out two rounds of reviews, and the authors were asked to address all the major and minor issues in their submissions during the review process. Eventually we accepted seven high-quality submissions in terms of clarity, novelty, significance, and relevance.

The first paper “COLIN: A Cache-Conscious Dynamic Learned Index with High Read/Write Performance” by Zhou Zhang *et al.* proposes a cache-conscious learned index that effectively supports insertions and addresses the problem of unbounded query complexity. The paper presents the structure of the index structure, the supported operations and an analysis of the model’s maximum error and the worst-case query cost.

The second paper “WATuning: A Workload-Aware Tuning System with Attention-Based Deep Reinforcement Learning” by Jia-Ke Ge *et al.* introduces a workload-aware configuration tuning system based on attention-based deep reinforcement learning. The paper designs a mechanism to classify workloads and then develops an effective deep reinforcement learning algorithm to make the recommended configurations more suitable for different workloads.

The third paper “Cardinality Estimator: Processing SQL with a Vertical Scanning Convolutional Neural Network” by Shao-Jie Qiao *et al.* proposes a vertical scanning convolutional neural network for cardinality estimation. The proposed approach first converts an SQL query into a feature map by considering semantic information including tables, joins, and predicates and then trains the proposed neural network to produce accurate estimation results.

The fourth paper “TransGPerf: Exploiting Transfer Learning for Modeling Distributed Graph Computation Performance” by Song-Jie Niu and Shi-Min Chen solves the scarcity of training examples for modeling the performance of distributed graph computation. The paper introduces a transfer learning based solution that exploits prior knowledge from a source scenario and utilizes a manageable amount of training data for modeling the performance of a target graph.

The fifth paper “Efficient Model Store and Reuse in an OLML Database System” by Jian-Wei Cui *et al.* introduces an OnLine Machine Learning (OLML) database system that stores and reuses trained machine learning models for new tasks. The paper develops an efficient reuse algorithm and evaluates the algorithm in two types of natural language processing (NLP) tasks.

The sixth paper “Impacts of Dirty Data on Classification and Clustering Models: An Experimental Evaluation” by Zhi-Xin Qi *et al.* empirically studies the impact of dirty data on two representative machine learning tasks, classification and clustering. The paper presents the evaluation methodology and provides an in-depth experimental analysis based on the evaluation.

The seventh paper “Mixed Hierarchical Networks for Deep Entity Matching” by Chen-Chen Sun and De-Rong Shen proposes mixed hierarchical deep neural networks for entity matching. The proposed approach exploits semantics from different abstract levels in the record internal hierarchy and then utilizes a family of attention mechanisms and a domain adaptation strategy to improve the performance.

### Leading Editor

Guo-Liang Li, Professor, Department of Computer Science and Technology, Tsinghua University, Beijing  
liguoliang@tsinghua.edu.cn

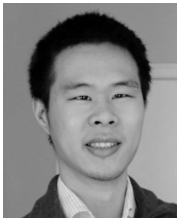
### Guest Editors

Dong Deng, Assistant Professor, Department of Computer Science, Rutgers University, Piscataway  
dong.deng@rutgers.edu

Ju Fan, Associate Professor, Renmin University of China, Beijing fanj@ruc.edu.cn



**Guo-Liang Li** is a full professor at Department of Computer Science and Technology, Tsinghua University, Beijing. He got his Bachelor’s degree in computer science from Harbin Institute of Technology, Harbin, in 2004, and Ph.D. degree in computer science from Tsinghua University, Beijing, in 2009. His research interests include AI4DB, DB4AI, big data management, crowdsourced data management, and large-scale data cleaning and integration. He got VLDB 2017 Early Research Contribution Award, TCDE 2014 Early Career Award, CIKM 2017 Best Paper Award, VLDB 2020 Best Papers, KDD 2018 Best Papers, ICDE 2018 Best Papers, DASFAA 2014 Best Paper Runnerup, APWeb 2014 Best Paper Award, EDBT 2013 Similarity Join and Search Champion. He was the general chair of SIGMOD 2021 and regularly served as a PC member of SIGMOD, VLDB, ICDE, KDD, WWW. He is serving as associate editor for IEEE TKDE, VLDB Journal, etc..



**Dong Deng** is an assistant professor in the Department of Computer Science at Rutgers University, Piscataway. His research interests include data management, database system, data curation, and data integration, with a focus on developing end-to-end systems and proposing scalable algorithms to manage real-world data. Before joining Rutgers, he was a postdoctoral associate in the Database Group at MIT CSAIL, where he worked with Mike Stonebraker and Sam Madden on data curation systems. He received his Ph.D. degree from Tsinghua University, Beijing, with the Highest Dissertation Award. He is a recipient of the prestige Siebel Scholarship, Microsoft PhD Fellowship, and Google PhD Fellowship and has been regularly publishing in top venues including SIGMOD, PVLDB, and ICDE.



**Ju Fan** is an associate professor at Renmin University of China, Beijing. He received his Ph.D. degree from Tsinghua University, Beijing, in 2012, and worked as a research fellow at School of Computing, National University of Singapore, Singapore, from 2012 to 2015. His research interests are in general area of data management, with emphasis on the topics including AI4DB and human-in-the-loop data preparation. He has published more than 40 papers at top conferences/journals, including SIGMOD, VLDB, ICDE and TKDE. He regularly served as a PC member for SIGMOD, VLDB and ICDE. He is also a recipient of ACM China Rising Star Award.