

2026 IEEE 42nd International Conference on Data Engineering Workshops (ICDEW 2026)

**Montreal, QC, Canada
4-8 May 2026**

IEEE Catalog Number:
ISBN:

CFP2645A-POD
979-8-3195-0576-7



Copyright © 2026, IEEE

All Rights Reserved

Copyright and Reprint Permissions:

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

*** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.

IEEE Catalog Number:	CFP2645A-POD
ISBN (Print-On-Demand):	979-8-3195-0576-7
ISBN (Online):	979-8-3195-0575-0
ISSN:	1943-2895

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA

Phone:	(845) 758-0400
Fax:	(845) 758-2633
E-mail:	curran@proceedings.com
Web:	www.proceedings.com

TABLE OF CONTENTS

A Divergent Index Tuning Advisor Using Quantum Machine Learning for Distributed Databases	1
<i>Sam Bird, Le Gruenwald, Sven Groppe</i>	
Rethinking Relational Operators as Hardware-Accelerated Matrix Operations	8
<i>Jannis Karampetsos, Mareike Schmidt, Martin Poppinga, Annett Ungethüm</i>	
Hot-Page-Aware Checkpointing for Flash SSDs	15
<i>Geunhyun Park, Sang-Won Lee</i>	
Efficient Quantification of Responsibility for the Spread of Misinformation and Disinformation Using the Banzhaf Index	22
<i>Masatoshi Yoshikawa, Hayate Asazawa</i>	
RewardHackingAgents: Benchmarking Evaluation Integrity for LLM ML-Engineering Agents	27
<i>Yonas Atinafu, Robin Cohen</i>	
Improving LLM performance through black-box online tuning: a case for adding system specs to Factsheets for Trusted AI	35
<i>Yonas Atinafu, Henry Lin, Robin Cohen</i>	
Trusting the Detector: Comparing Explanations for Deepfake Identification Across Modalities	43
<i>Lucas Kopp, Alina Lytovchenko, Robin Cohen</i>	
Structure Selection for Fairness-Constrained Differentially Private Data Synthesis	53
<i>Naeim Ghahramanpour, Mostafa Milani</i>	
From Water Networks to Wellness: Data Science–Driven Design of Human-Inhabited Spaces	62
<i>Genoveva Vargas-Solar, Peace Chinemere Ebika, José-Luis Zechinelli Martini, Celeste Cedillo Gonzalez Javier A. Espinosa-Oviedo</i>	
The Human Factor in Data Cleaning: Exploring Preferences and Biases	69
<i>Hazim AbdElazim, Shadman Islam, Mostafa Milani</i>	
Artificial intelligence and Women’s Work: a Review of evidence	78
<i>Francesca Zafonte, Tania Cerquitelli</i>	
Improving Sentiment Analysis of Mobile App Reviews through Emoji and Slang Normalization	85
<i>Tinne Dey, Zheyang Zhang, Kostas Stefanidis</i>	
Second International Workshop on LLM+Vector Data: Agentic RAG Edition	93
<i>Arijit Khan, Yuyu Luo, M. Tamer Özsu, Danrui Qi, Jiannan Wang</i>	
VecAlign: Efficient Vector Sequence Alignment for Similarity Search in the Embedding Space	97
<i>Zhizhi Wang, Dong Deng</i>	
Performance of Small Language Model Pretraining on FABRIC: An Empirical Study	111
<i>Praveen Rao</i>	
CORE - An LLM-Powered System for Resolution Mining and Retrieval from Conversational Platforms	123
<i>Archan Dutta, Vinay Raj Sisodiya, Hardik Airen, Phani Nivarthi</i>	

Cost Trade-offs of Reasoning and Non-Reasoning Large Language Models in Text-to-SQL	134
<i>Saurabh Deochake, Debajyoti Mukhopadhyay</i>	
Automatic End-to-End Data Integration using Large Language Models	142
<i>Aaron Steiner, Christian Bizer</i>	
Towards Executing Sloppy SQL Queries Over Tabular Data Lakes	151
<i>Jan-Micha Bodensohn, Jakob Steinke, Carsten Binnig</i>	
How Far Can They Map? Probing LLM Capabilities for Cross-Schema SQL Generation	154
<i>Mohammadreza Daviran, Davood Rafiei</i>	
TiRex: An HTAIP Framework Beyond HTAP for Unified Transactional, Analytical, and AI Workloads	159
<i>Jane Yu, Yu Dong, Rossi Sun, Lucas Sun, Liu Tang, Ed Huang, Max Liu</i>	
Less Is More? When Dataset Context Hurts LLM-Generated Dataset Descriptions	166
<i>Lisa-Yao Gan, Arunav Das, Johanna Walker, Klaus Diepold, Elena Simperl</i>	
LLM-generated Schema Matching with Formally Verifiable Explanations	176
<i>Dante Pinto, Len Feremans, Frank Neven, Marcel Parciak, Stijn Vansummeren</i>	
A2RAG: Adaptive Agentic Graph Retrieval for Cost-Aware and Reliable Reasoning	187
<i>Jiate Liu, Zebin Chen, Shaobo Qiao, Mingchen Ju, Danting Zhang, Bocheng Han, Shuyue Yu, Xin Shu Jinglin Wu, Dong Wen, Xin Cao, Guanfeng Liu, Zhengyi Yang</i>	
Implicit to Explicit: A Pipeline for Inter-Policy Conflict Detection in Data Ecosystems Governance	197
<i>Achraf Hmimou, Petar Jovanovic, Oscar Romero</i>	
Prompt Optimization for User Simulation in Conversational Recommender Systems: A Multi-Objective Framework	204
<i>Nipun B Nair, Tongtong Wu, Weiqing Wang</i>	
A Cloud-Edge Architecture for Automated Metadata Generation in Smart Environments	214
<i>Muhammad Ramish, Ahmed Khalid, Evan Center, Ekaterina Gilman, Lauri Loven, Hassan Mehmood</i>	
CityChat: Template-Guided Text-to-SQL over City-Scale Spatial Data	222
<i>Yiyang Bian, Amr Magdy</i>	
Open-Data Traffic Noise Prediction across Cities and Traffic Stations	231
<i>Danila Goncharenko, Amirhossein Ghaffari, Ekaterina Gilman</i>	
Semi-Supervised Object Detection with Noisy Urban Street Image Datasets for Smart City	240
<i>Bhavyesh Sajja, Seon Ho Kim</i>	
How Private Are Genomic Embeddings? Assessing and Mitigating Reconstruction Attacks on Transformer-Based DNA Language Models Embeddings	248
<i>Reem Al-Saidi</i>	
Fed-TTC: A Pareto-Efficient Hierarchical Federated Learning Framework	253
<i>Seyed Salar Ghazi</i>	
Private Set Operations Using Doubly Oblivious Radix Partitioning	258
<i>Idil Kara</i>	
Semantic-Driven Temporal Order Reconstruction	263
<i>Wongyu Kim</i>	