

33rd Italian Symposium on Advanced Database Systems (SEBD 2025)

Ischia, Italy
16-19 June 2025

ISBN: 979-8-3313-3333-1

Printed from e-media with permission by:

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



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© for individual papers are retained by the authors.
All rights reserved.

Printed by Curran Associates, Inc. (2026)

Edited by:
Ilaria Bartolini
Gianluca Cima
Donatella Firmani
Domenico Lembo
Davide Martinenghi
Massimo Mecella
Antonella Poggi

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
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

- [Preface](#)
Summary: There were 58 peer-reviewed papers. Additionally, 8 papers were accepted to be presented at the Doctoral Consortium and 7 papers were presented as a poster presentation.

Keynotes

- [A Theory of Regular Queries - Abstract](#) 1–1
Moshe Y. Vardi
- [Local Consistency vs. Global Consistency on Annotated Data - Abstract](#) 2–2
Phokion G. Kolaitis
- [SQL and Large Language Models: A Marriage Made in Heaven? - Abstract](#) 3–3
Paolo Papotti

Session 1: Privacy and Security

- [Privacy of Textual Data: The pyPANTERA Package](#) 4–14
Francesco Luigi De Faveri, Guglielmo Faggioli, Nicola Ferro
- [QulPU: Evaluating Actual Privacy of Obfuscated Queries](#) 15–26
Francesco Luigi De Faveri, Guglielmo Faggioli, Nicola Ferro
- [Password Strength Analysis Through Social Network Data Exposure: A Combined Approach Relying on Data Reconstruction and Generative Models](#) 27–35
Maurizio Atzori, Eleonora Calò, Loredana Caruccio, Stefano Cirillo, Giuseppe Polese, Giandomenico Solimando
- [Privacy-Preserving Identity Management in Cloud Environments via SOLID](#) 36–46
Alfredo Cuzzocrea, Islam Belmerabet, Ismail Benlaredj

Session 2: Knowledge Representation

- [On Equality Constraints in Datalog+/- Knowledge Bases](#) 47–55
Andrea Cali, Marco Console, Riccardo Frosini
- [Data Augmentation for Data-Centric AI Through the Lens of Semantic Technologies: A Position Paper](#) 56–66
Daniele Bertillo, Luca Cabibbo, Gianluca Cima, Valter Crescenzi, Marco Console, Roberto Maria Delfino, Stefano Iannucci, Domenico Lembo, Maurizio Lenzerini, Lorenzo Marconi, Paolo Merialdo, Marco Napoleone, Laura Papi, Antonella Poggi, Federico Maria Scafoglieri, Riccardo Tortone
- [Semantic Containment in MLMs: A Prompt-Based Approach](#) 67–79
Vito Walter Anelli, Alessandro De Bellis, Tommaso Di Noia, Eugenio Di Sciascio
- [Enhancing Data Management and Value Creation through a Knowledge-centric Data Stack](#) 80–94
Enrico Franconi, Théo Abgrall
- [Towards Leveraging Link Prediction for Geospatial Data Integration](#) 95–104
Albulen Pano, Mattia Fumagalli, Davide Lanti, Diego Calvanese
- [MARTIALIS: An Open Framework for Knowledge Graphs-based Retrieval Augmented Generation](#) 105–114
Edoardo Bianchini, Filippo Bianchini, Marco Calamo, Francesca De luzi, Mattia Macrì, Massimo Mecella,

- [The Vadalog Parallel System: Distributed Reasoning with Datalog+/- \(extended abstract\)](#) 115–128
Luigi Bellomarini, Davide Benedetto, Matteo Brandetti, Emanuel Sallinger, Adriano Vlad

Session 3: Social Networks, Data Mining and Information Retrieval

- [Relevant, yet Hard to Find: Directional Queries to the Rescue](#) 129–138
Paolo Ciaccia, Davide Martinenghi
- [Benchmarking Active Learning Techniques: Insights from Multi-Domain Fake News Detection](#) 139–154
Sergio Flesca, Marco Gagliardi, Danilo Maurmo, Francesco Scala, Eugenio Vocaturo
- [Analyzing the dynamics of user influence in Threads](#) 155–164
Gianluca Bonifazi, Christopher Buratti, Enrico Corradini, Michele Marchetti, Federica Parlapiano, Davide Traini, Domenico Ursino, Luca Virgili
- [Joint Information Retrieval and Recommendation: a Reproducibility Study](#) 165–175
Simone Merlo, Guglielmo Faggioli, Nicola Ferro
- [Integrating Semantic, Social, and Spatial Dimensions for Inductive Malicious User Detection in Social Networks](#) 176–184
Francesco Benedetti, Antonio Pellicani, Gianvito Pio, Michelangelo Ceci
- [A Dimension Importance Estimation-based Framework for Query Performance Prediction](#) 185–195
Guglielmo Faggioli, Nicola Ferro, Raffaele Perego, Nicola Tonellotto
- [Concept Drift Detection in Machine Learning Systems by Exploiting Relaxed Functional Dependencies](#) 196–205
Loredana Caruccio, Stefano Cirillo, Giuseppe Polese, Roberto Stanzone
- [Discovering Balance-Aware Polarized Communities in Signed Networks with Graph Neural Networks](#) 206–216
Francesco Gullo, Domenico Mandaglio, Andrea Tagarelli

Session 4: Cloud and IoT

- [Design of Cloud Data Platforms](#) 217–228
Alex Baiardi, Matteo Francia, Enrico Gallinucci, Matteo Golfarelli, Manuele Pasini
- [A Unified Framework for Human Activity Recognition Data Preprocessing](#) 229–237
Sara Comai, Davide Mangano, Kévin Bouchard, Fabio Salice
- [A Process-Aware Model for Industrial IoT Integration and Analytics based on Knowledge Graphs](#) 238–246
Claudia Diamantini, Alex Mircoli, Domenico Potena, Emanuele Storti
- [Towards Bringing Human-Like Intelligence in the Computing Continuum: the INTEND Project](#) 247–255
Lorenzo Balzotti, Donatella Firmani, Francesco Leotta, Jerin George Mathew, Jacopo Rossi
- [Leveraging LLMs and RAG for Data Service Discovery in the Internet of Production](#) 256–263
Devis Bianchini, Valeria De Antonellis, Massimiliano Garda, Michele Melchiori, Anisa Rula

- [LLM-Based Composition of Smart Data Services in Shop Floors](#) 264–273
Adriano Izzi, Jerin George Mathew, Flavia Monti, Donatella Firmani, Francesco Leotta, Federica Mandreoli, Massimo Mecella
- [A Framework for Enhanced Decision-Making Support in Production Planning through Event Data Analysis](#) 274–283
Simone Agostinelli, Dario Benvenuti, Angelo Casciani, Francesca De Luzi, Matteo Marinacci, Andrea Marrella, Jacopo Rossi
- [Supporting Multidimensional Risk Analysis and Mining over Big Digital Twins: An Effective and Efficient Framework](#) 284–299
Alfredo Cuzzocrea, Ismail Benlaredj

Session 5: Data-Driven Decision Making

- [AI-Driven Clinical Reporting: A Case Study on IQLINIQ](#) 300–316
Seyedeh Leili Mirtaheri, Reza Shahbazian, Narges Movahedkor, Irina Trubitsyna, Sergio Greco
- [Comparative Conflict Analysis between Autonomous and Human-Operated Vehicles with Pedestrians at Unsignalized Crosswalks](#) 317–326
Andrea Avignone, Marco Bassani, Beatrice Borgogno, Brunella Caroleo, Silvia Chiusano, Federico Princiotta
- [Processing and Representation of Linguistic Properties in Large Language Models](#) 327–336
Elisabetta Rocchetti, Alfio Ferrara
- [Analyzing Common Student Errors in SQL Query Formulation to Enhance Learning Support](#) 337–345
Davide Ponzini, Abdolhamid Livani, Giovanna Guerrini, Barbara Catania, Mauro Coccoli
- [Legal Document Query Language: Conceptualizing Linguistic Commands for AI Assistants in Civil Appeal Proceedings](#) 346–359
Ruben Agazzi, Carlo Batini, Matteo Palmonari, Monica Vitali

Session 6: Data Quality, Curation and Fairness

- [Binomial Confidence Intervals for Knowledge Graph Accuracy Estimation \(Extended Abstract\)](#) 360–369
Stefano Marchesin, Gianmaria Silvello
- [BUNNI: Discover Repair Actions for Data Cleaning](#) 370–381
Giansalvatore Mecca, Paolo Papotti, Donatello Santoro, Enzo Veltri
- [PuckTrick: A Library for Making Synthetic Data More Realistic](#) 382–398
Alessandra Agostini, Andrea Maurino, Blerina Spahiu
- [On Measuring Inconsistency in Graph Databases](#) 399–410
John Grant, Francesco Parisi
- [Analysing the Impact of Data Distribution Shifts on Model Fairness in Machine Learning](#) 411–422
Federico Motta, Yijie Li, Huiping Chen, Federica Mandreoli, Paolo Missier
- [S-PIC4CHU: Semantics-based Provenance, Integrity, and Curation for Consistent, High-quality, and Unbiased Data Science](#) 423–433
Gianvincenzo Alfano, Ilaria Bartolini, Diego Calvanese, Paolo Ciaccia, Sergio Greco,

Davide Lanti, Emilia Lenzi, Davide Martinenghi, Cristian Molinaro, Marco Patella, Letizia Tanca, Riccardo Torlone, Irina Trubitsyna

- [Divergence-aware Approaches to Mitigate Subgroup Disparities in Speech Models](#) 434–445

Alkis Koudounas, Eliana Pastor, Flavio Giobergia, Elena Baralis,

- [An Associative Approach to Fair Co-clustering](#) 446–454

Federico Peiretti, Ruggero Pensa

Session 7: Applications

- [Extraction and Investigation of Power Neurons in the *Caenorhabditis elegans* Connectome](#) 455–469

Christopher Buratti, Michele Marchetti, Federica Parlapiano, Giorgio Terracina, Domenico Ursino

- [A Deep Learning-Based Framework for Stock Analysis](#) 470–477

Fausto Ricchiuti, Giancarlo Sperlì

- [Optimizing Defect Detection: A Machine Learning-Ready Data Processing Pipeline](#) 478–486

Loredana Cristaldi, Emilia Lenzi, Davide Martinenghi, Luca Martiri, Andrea Moschetti, Letizia Tanca, Marco Zanoni

- [The Role of Artificial Intelligence in ESG Ranking Evaluation](#) 487–494

Simona Fioretto, Enea Vincenzo Napolitano, Elio Masciari

- [Metadata as a Key Driver in Healthcare Data Analysis](#) 495–505

Chiara Criscuolo, Davide Piantella, Pierluigi Reali, Maria Gabriella Signorini, Letizia Tanca

- [A Digital Forensics Approach for Validating Screen Captures in Legal Proceedings](#) 506–515

Claudia Greco, Michele Ianni, Antonella Guzzo, Giancarlo Fortino

- [Leveraging Ontologies to Enhance Personalised and Sustainable Food Recommendations with LLMs](#) 516–524

Ada Bagozi, Devis Bianchini, Michele Melchiori, Anisa Rula

- [MiCare: An IoT-Based System for Real-Time Mental Health Monitoring and Early Disease Detection](#) 525–540

Marco Cremaschi, Sara Nocco, Alessandra Agostini, Andrea Maurino

Session 8: Interpretable Data-centric AI

- [Explanations for Negative Query Answers under Inconsistency-Tolerant Semantics](#) 541–550

Thomas Lukasiewicz, Enrico Malizia, Cristian Molinaro

- [Using Heterogeneous Graph Neural Networks for Explainable Next-Activity Prediction](#) 551–559

Vincenzo Pasquadibisceglie, Raffaele Scaringi, Annalisa Appice, Giovanna Castellano, Donato Malerba

- [From XAI to XEE: eXplainable End-to-End using Influence and Provenance](#) 560–569

Paolo Missier, Riccardo Torlone, Pasquale Leonardo Lazzaro, Luca Gregori, Clyde Fernandes, Tianyang Xie

- [Explaining Entity Matching Models with CREW](#) 570–579
Riccardo Benassi, Michele Luca Contalbo, Francesco Del Buono, Francesco Guerra, Giacomo Guiduzzi, Matteo Paganelli, Sara Pederzoli, Donato Tiano, Maurizio Vincini
- [Supporting the Analysis and Explanation of Dependencies through an LLM-Powered Tool](#) 580–593
Loredana Caruccio, Tullio Pizzuti, Giuseppe Polese

Session 9: Natural Language and Text Processing

- [JusBuild: a RAG-based Architecture for Legal Document Building - Discussion paper](#) 594–603
Silvana Castano, Alfio Ferrara, Stefano Montanelli, Sergio Picascia, Davide Riva
- [Managing Legal Data: A Framework for Document Annotation and Retrieval](#) 604–614
Valerio Bellandi, Stefano Siccardi
- [In-Context LLM Query Execution on Textual Documents](#) 615–625
Dario Satriani, Enzo Veltri, Donatello Santoro, Paolo Papotti
- [WhosAI: A Contrastive Learning Framework for Machine-Generated Text Detection and Attribution](#) 626–636
Lucio La Cava, Andrea Tagarelli
- [Text-to-Refused-SQL: A Comprehensive Evaluation of LLMs Refusal in Text-to-SQL](#) 637–651
Giovanni Sullutrone, Luca Sala, Lisa Trigiante, Sonia Bergamaschi

Doctoral Consortium

- [A Symbolic/Sub-Symbolic-Based Framework for Indicators Management](#) 652–658
Cristina Rossetti
- [Preserving Privacy When Processing User Queries in IR](#) 659–666
Francesco Luigi De Faveri
- [Towards Joint Information Retrieval and Recommender Systems](#) 667–674
Simone Merlo
- [Design and Development of a Polystore System for Heterogeneous Biomedical Data](#) 675–681
Mirco Cazzaro
- [Advancing Cross-Document Relation Extraction with Hybrid Retrieval and Knowledge-Augmented Reasoning](#) 682–688
Marco Martinelli
- [Towards the Definition of a Unified Metamodeling Semantics for Knowledge Graphs](#) 689–694
Roberto Maria Delfino
- [Semantic Explanations and Reasoning for Machine Learning Models: An Ontology-Based Approach](#) 695–700
Laura Papi
- [Querying OpenEO Datacubes through Virtual Knowledge Graphs](#) 701–707
Albulen Pano

Research Projects Exhibition

- [Research Project Exhibition Track SEBD 2025](#) 708–724
Francesca De Luzi, Flavia Monti, Massimo Mecella