

Giovanni Simonini

Associate Professor – Information Processing Systems (IINF-05/A)
University of Modena and Reggio Emilia, Dept. of Engineering “Enzo Ferrari”
simonini@unimore.it · www.giovanisimonini.com
Updated: September 2026

Research Profile

Giovanni Simonini is Associate Professor at the University of Modena and Reggio Emilia, working on scalable data management, data integration, and AI-driven methods for industrial data analysis. His current research includes **LLM context management** and the application of **agentic LLM-based techniques** to industrial settings, with a focus on reliable data-centric AI systems, process support, automation, and human-in-the-loop data management.

Current and Previous Positions

Associate Professor Nov 2022 – Present

University of Modena and Reggio Emilia, Dept. of Engineering “Enzo Ferrari”

Research area: Scalable Data Management; LLM context management. Qualified for Full Professorship, Italian National Scientific Qualification, IINF-05/A, since Nov. 19, 2024.

Assistant Professor, RTD-B Nov. 2019 – Oct 2022

University of Modena and Reggio Emilia

Research on scalable data management.

Postdoctoral Associate Sept. 2018 – Nov. 2019

Massachusetts Institute of Technology – Computer Science and Artificial Intelligence Laboratory

Research with Prof. Michael Stonebraker on automated data cleaning and preparation in machine learning pipelines.

Research Fellow Jan 2016 – Sep 2018

University of Modena and Reggio Emilia

Research on scalable data integration and entity resolution.

PhD Student Jan 2013 – Jan 2016

University of Modena and Reggio Emilia

PhD research on data integration and entity resolution. Co-advisor: Prof. H. V. Jagadish, University of Michigan.

Selected Scientific and Industrial Research Projects

Industrial Collaboration with SPAL Automotive 2026 – Present

Role: Scientific Coordinator. Active collaboration on the application of **AI and LLM-based agentic techniques** in industrial settings, including the supervision of two PhD positions focused on AI-driven industrial data analysis, process support, and automation.

AI for Industry – AIMAG Spa 2026

Industrial project on the application of AI methods to industrial data and processes.

Emilia-Romagna PR-FESR 2022 DATHA 2023 – 2026

Role: Principal Investigator. Regional research and innovation project on data-driven methods and technologies.

PRIN 2022 National Project – “Discount Quality for Responsible Data Science” 2023 – 2025

Role: Local Principal Investigator. MIUR-funded project on data quality and responsible data science.

ML Techniques for Wearable Devices 2021

Role: Principal Investigator. Industrial project on machine learning techniques for wearable devices.

ECOSISTER, PNRR 2025

Role: Participant. National Recovery and Resilience Plan project.

Core Competences

- **Agentic LLMs:** LLM-based agents for industrial data analysis, automation, and decision support.
- **LLM context management:** data preparation, retrieval, and integration for reliable LLM applications.
- **Scalable data management:** data integration, data preparation, data quality, and data cleaning at scale.
- **Vector and hybrid search:** adaptive hybrid search in vector databases and scalable indexing.
- **Industrial AI:** company collaborations on analytics, automation, AI workflows, and applied machine learning.

Teaching and PhD Supervision

- Big Data Management, MSc in Computer Engineering and AI Engineering, UNIMORE, 2023 – Present.
- Quantitative Methods and Computer Science, Data Analysis for Economics, UNIMORE, 2023 – Present.

Additional teaching

- Visiting Professor, Aalborg University, PhD course on Big Data Integration, 2024.
- Tools and Techniques for Big Data Analysis, CINECA, four editions, 2014 – 2017.

PhD supervision

- 2 ongoing PhD students as advisor: Bilel Arfaoui, supported by Skybackbone Engenio; Giovanni Malaguti, supported by Leonardo S.p.A.
- 2 completed PhD students as advisor: Adeel Aslam, now Postdoc at University of Southampton; Giulio de Sabbata, now Data Scientist.
- 1 completed PhD student as co-advisor: Luca Zecchini, now Researcher at TU Berlin.
- 2 PhD positions within the active SPAL Automotive collaboration on AI and LLM-based agentic techniques for industrial applications.

Selected Collaborations and Networks

Felix Naumann, Hasso Plattner Institute, Germany, on entity resolution, data integration, and table overlap; Michael Stonebraker, MIT CSAIL, USA, on automated data cleaning and preparation in ML pipelines; H. V. Jagadish, University of Michigan, USA, on data integration and entity resolution; Themis Palpanas, Université Paris Cité, France, on scalable entity resolution and meta-blocking; George Papadakis, University of Athens / Athena RC, Greece, on entity resolution and meta-blocking.

Awards and Service

- National Scientific Qualification for Full Professor, IINF-05/A, 2024.
- Honorable Mention for AI and Big Data, Gruppo 2003, 2021.
- Best PhD Thesis Award, IEEE Computer Society – Italy Section, 2017.
- Curriculum Coordinator, Computer Science and Engineering, PhD in ICT, UNIMORE, 2024 – Present.
- Faculty Board Member, PhD in ICT, UNIMORE, 2019 – Present.
- Proceedings Chair, VLDB 2025 and 2026.
- Co-Chair, ACM SIGMOD Programming Contest, 2021 and 2022.
- Program Committees (multiple editions): ICDE, SIGMOD, VLDB, EDBT.
- Reviewer for IEEE TKDE, Data & Knowledge Engineering, ACM JDIQ.

Selected Publications

1. A. Aslam, L. Gagliardelli, E. Rezig, G. Konstantinidis, G. Simonini. Versatile Sketch-Based Attribute Filtering for Hybrid Vector Search. *IEEE Transactions on Knowledge and Data Engineering*, pp. 6270-6284, vol. 38, 2026.
2. A. Aslam, R. Khan, G. Simonini, and G. Konstantinidis. Adaptive query-aware hybrid search in vector databases. *EDBT*, 2026.
3. L. Zecchini, V. Efthymiou, F. Naumann, and G. Simonini. Deduplicated sampling on-demand. *Proceedings of the VLDB Endowment*, 2025.
4. L. Zecchini, Z. Abedjan, V. Efthymiou, and G. Simonini. RadLER: Deduplicated sampling on-demand. *Proceedings of the VLDB Endowment*, 2025.
5. B. Pernici et al. Sustainable quality in data preparation. *ACM Journal of Data and Information Quality*, 2025.
6. L. Zecchini, T. Bleifuß, G. Simonini, S. Bergamaschi, and F. Naumann. Determining the largest overlap between tables. *Proceedings of the ACM on Management of Data*, 2024.
7. G. Simonini, L. Zecchini, S. Bergamaschi, and F. Naumann. Entity resolution on-demand. *Proceedings of the VLDB Endowment*, 2022.
8. L. Gagliardelli, G. Papadakis, G. Simonini, S. Bergamaschi, and T. Palpanas. Generalized supervised meta-blocking. *Proceedings of the VLDB Endowment*, 2022.
9. E. K. Rezig, L. Cao, G. Simonini, M. Schoemans, S. Madden, N. Tang, M. Ouzzani, and M. Stonebraker. Dagger: A data, not code, debugger. *CIDR*, 2020.
10. E. K. Rezig et al. Data Civilizer 2.0: A holistic framework for data preparation and analytics. *Proceedings of the VLDB Endowment*, 2019.
11. G. Simonini, G. Papadakis, T. Palpanas, and S. Bergamaschi. Schema-agnostic progressive entity resolution. *IEEE TKDE*, 2019.
12. G. Simonini, L. Gagliardelli, S. Bergamaschi, and H. V. Jagadish. Scaling entity resolution: A loosely schema-aware approach. *Information Systems*, 2019.
13. G. Simonini, S. Bergamaschi, and H. V. Jagadish. BLAST: A loosely schema-aware meta-blocking approach for entity resolution. *Proceedings of the VLDB Endowment*, 2016.