



Matteo Paganelli

Researcher

matteo.paganelli@unimore.it

[+393348264672](tel:+393348264672)

[Modena, Italy](#)

[linkedin.com/in/matteo-paganelli-71410a150](https://www.linkedin.com/in/matteo-paganelli-71410a150)

[matteo-paganelli.github.io](https://github.com/matteo-paganelli)

Summary

Matteo Paganelli is a tenure-track researcher at the "Marco Biagi" Department of Economics, University of Modena and Reggio Emilia. He previously conducted postdoctoral research at the Hasso Plattner Institute in Potsdam and received his Ph.D. in Information and Communication Technologies from the University of Modena and Reggio Emilia in 2021. He holds the Italian National Scientific Qualification for Associate Professor in Information Processing Systems and Informatics. His research focuses on data management and integration, language and tabular AI, and trustworthy AI, including explainability and fact-checking.

Professional Experience

12/2024 – Present	University of Modena and Reggio Emilia Tenure-track researcher (RTT)
05/2024 – 12/2024	University of Modena and Reggio Emilia Postdoc The research activity focused on analyzing corporate ESG performance through text analysis techniques, with a particular emphasis on the use of advanced language models (LLMs) for the automatic extraction of sustainability indicators and the detection of potential greenwashing practices.
05/2023 – 05/2024 Germany	Hasso Plattner Institute Postdoc The research activity focused on the implementation of robust entity matching models for noisy data, the development of algorithms for detecting duplicate tables, and the analysis of schema matching approaches based on functional dependencies.
01/2022 – 04/2023	Cristail (Startup of the University of Modena and Reggio Emilia) Machine Learning Engineer
01/2021 – 01/2022	University of Modena and Reggio Emilia Postdoc The research activity focused on the development of machine learning and deep learning models for anomaly detection in industrial data.
09/2020 – 12/2020 Internship completed remotely due to COVID-19	Utrecht University Research Intern The research activity focused on the unsupervised analysis of the quality of data integration results.

Education

11/2017 – 03/2021	PhD in Information and Communication Technologies (ICT) University of Modena and Reggio Emilia
09/2015 – 10/2017	Master's degree in Computer Engineering University of Modena and Reggio Emilia
09/2011 – 09/2015	Bachelor's degree in Computer Engineering University of Modena and Reggio Emilia

Languages

Italian



English



Certificates

Italian National Scientific Qualification for Associate Professor

In the call 2023/2025 (Ministerial Decree n. 1796/2023) for the disciplinary fields of Information Processing Systems (09/H1) and Informatics (01/B1).

Funded Research Projects

2024 – 2026

PANACEA (A Model-Based Framework for Self-Protecting Systems)

PRIN 2022 - Research Team Member

The project aims to develop an explainable network intrusion detection system (NIDS) capable of adapting to data-distribution shifts, as well as to create a public and standardized benchmark for the validation and comparison of self-protection technologies.

2024 – 2026

RESISTo (A Digital Twins-enabled platform for a REsilient and Sustainable production in the InduSTry 5.0 era)

Regional PR FESR 2021-2027 - Work Package Lead

The project focuses on the development of a large language model-based system for the generation of ESG indicators and the detection of ESG-washing practices through (semi-)automatic analysis of textual data.

2019 – 2021

Big Data In and Out Services for Industry 4.0 (SBDIO I4.0)

Regional POR FESR 2014-2020 - Research Team Member

2017 – 2019

Research Alps

EU CEF Telecom - Scientific Task Coordinator

The project aims to create an open dataset describing research laboratories located in the Alpine region by integrating existing open and heterogeneous data sources.

Industry and Institutional Projects

2026 – Present

Supplier Matching and Product Classification

IUNGO S.p.A. - Research consultancy

The project focuses on the development of a data integration pipeline to consolidate duplicate supplier records, classify heterogeneous product descriptions according to the UNSPSC taxonomy, and build standardized supplier profiles.

2025 – Present

Counterfactual Explanations for Credit Scoring

CRIF S.p.A. - Principal Investigator

The project focuses on the development of counterfactual explanation techniques for machine learning models applied to credit scoring, aiming to enhance the transparency and interpretability of predictive systems.

2020

Advanced Sales Data Analysis and Production Scheduling

Italgraniti S.p.A. - Research consultancy

2020

Spatial and Temporal Open Data Platform

Autorità di Bacino Distrettuale del Fiume Po - Research consultancy

2019

Time-Series Analysis and Performance Indicators

Ceramiche Polis S.p.A. - Research consultancy

Academic service

2019 – Present

Conference Presentations

Presented peer-reviewed research at EACL 2026, ADBIS 2026, ICDE 2024, VLDB 2022, SIGMOD 2021, EDBT 2021, iiWAS 2020, CIKM 2019, ADBIS 2019, and IEEE Big Data Congress 2019.

2018 – Present

Academic Reviewing and Program Committees

Program Committee: SIGMOD (2023, 2025-2026), CIKM (2024, 2026), DOLAP (2024-2025), EDBT (2027), PVLDB (2027).

Reviewer: ACL Rolling Review (2026), VLDB Journal (2024), Journal of Data and Information Quality (2025), Journal on Data Semantics (2021), Multimedia Tools and Applications (2023), CIKM (2018, 2020-2021), WSDM (2021), and SIGIR (2021).

Awards

Distinguished Reviewer Award for SIGMOD 2025.

Academic Teaching

2025 – Present

Business Intelligence

Course Instructor

MSc in International Management at the "Marco Biagi" Department of Economics, University of Modena and Reggio Emilia (6 ECTS - ING-INF/05).

2025 – Present

Human Resource Information Systems and Data Science

Course Instructor

MSs in Labour Relations at the "Marco Biagi" Department of Economics, University of Modena and Reggio Emilia (6 ECTS - ING-INF/05).

2025

Computer Science (Basic Qualification)

Course Instructor

BSc in Economics and Finance at the "Marco Biagi" Department of Economics, University of Modena and Reggio Emilia (3 ECTS - INF/01).

2024

Exploring the Frontier of Database Research: the Alliance with Foundation Models

Course Instructor

12-hour module within the ICT Doctoral School at the University of Modena and Reggio Emilia (3 ECTS).

2024

Constraint-based Schema Matching

Co-Instructor

MSc in Data Engineering at the Hasso Plattner Institute / University of Potsdam (12 ECTS).

2017 – 2020

Big Data Analytics

Teaching Assistant

MSc in Computer Engineering at the University of Modena and Reggio Emilia (9 ECTS).

2017 – 2020

Software Engineering

Teaching Assistant

BSc in Computer Engineering at the University of Modena and Reggio Emilia (9 ECTS).

Research Profile

Data Integration and Matching

This research area addresses the integration of heterogeneous data at the record, schema, and table levels. It includes methods for configuring rule-based Entity Matching systems through TunER [c2], integrating noisy and heterogeneous data sources in real-world settings [c1, c4], and combining transformer-based representations with Automated Machine Learning for entity deduplication [c10].

A further component concerns the analysis of transformer-based Entity Matching models, with particular attention to their behaviour, performance, and sensitivity to data characteristics [j3, j6, c20]. Additional contributions cover synonymous attribute identification [c5], privacy-preserving schema matching based on functional dependencies [c22], and the unsupervised evaluation of data integration results in the absence of ground-truth data [c6, c14].

At the table level, Armadillo [j7] uses graph-based representations to estimate overlap between tables, supporting the identification of duplicate, related, or evolving tables in large collections.

Explainable and Trustworthy AI

This research area focuses on the transparency, interpretability, and actionability of machine learning systems. A substantial part of the work concerns Entity Matching models, for which explanations identify the attributes, words, or word combinations that influence matching decisions.

The proposed approaches include post-hoc explanation methods based on input perturbations and landmarks [c9, c11, c13], intrinsically interpretable architectures such as WYM [c15, c17], and methods that organize influential terms into compact and understandable groups [c19, c24]. These contributions investigate the trade-off between explanation fidelity, interpretability, and usability.

The same research direction extends to decision-support systems in other domains. CREDIX studies counterfactual explanations constrained by domain expertise, with the objective of generating realistic and actionable alternatives for credit-scoring decisions. Privacy-aware data integration [c22] and transparent analysis of corporate information [c25] provide further connections with trustworthy data-centric AI.

Language Models for Tabular and Document Intelligence

This research area investigates the use of language models for understanding, querying, and reasoning over structured and semi-structured information.

GRI-QA [c23] provides a benchmark for question answering over complex environmental tables extracted from corporate sustainability reports. The benchmark covers numerical reasoning, heterogeneous table structures, and questions requiring the combination of information distributed across multiple tables and documents.

CLARIESG [c25] provides an end-to-end framework for analysing ESG information in corporate reports. The system combines table extraction, structured prompting, unit normalization, table selection, and program-based reasoning to support transparent and verifiable analyses of sustainability indicators and potential ESG-washing practices.

Related work addresses natural language understanding and educational applications, including argument-relation classification through discourse information and adversarial training [c21], and the use of controllable multi-agent language-model architectures for guided reading comprehension [c26].

In-DBMS Machine Learning Inference

This research area studies the execution of machine learning pipelines directly within relational database management systems. The main objective is to reduce data movement and simplify deployment by translating preprocessing operations and predictive models into SQL.

MASQ [c7] introduced the transformation of machine learning pipelines into SQL queries. The corresponding systematic evaluation [j5] analyses relational databases as platforms for prediction serving, considering execution efficiency, deployment complexity, and compatibility with different models and preprocessing operations. An extended presentation of this work is provided in [c18].

Additional Research Contributions

Additional contributions concern the interactive exploration and automatic description of large datasets [c3, c8, j2], novelty and anomaly detection in industrial and network environments [c12, j4, j8], quantitative machine learning for financial applications [c16], and parallel algorithms for relational data integration [j1].

Complete Publication List

2026

- [c26] Chiara Magurno, Michele Luca Contalbo, Matteo Paganelli, Enrico Giliberti, Chiara Bertolini, Francesco Guerra: **CLAIRE: A Controllable LLM Tutoring Framework for Reading Comprehension**. AIED 2026: 49–63.
- [c25] Marta Santacroce, Michele Luca Contalbo, Sara Pederzoli, Riccardo Benassi, Valeria Venturelli, Matteo Paganelli, Francesco Guerra: **CLARIESG: An End-to-End System for ESG Analysis over Complex Tables in Corporate Reports**. EACL (System Demonstrations) 2026: 86–100.

2025

- [j8] Sara Pederzoli, Matteo Paganelli, Michele Luca Contalbo, Riccardo Benassi, Donato Tiano, Stefano Iannucci, Francesco Guerra: **NOCTOWL: Adaptive Tree-Based Model for Network Anomaly Detection Under Delayed and Sampled Label Availability**. IEEE Access 13: 197899–197911 (2025).
- [j7] Francesco Pugnalone, Luca Zecchini, Matteo Paganelli, Matteo Lissandrini, Felix Naumann, Giovanni Simonini: **Table Overlap Estimation through Graph Embeddings**. Proc. ACM Manag. Data 3(3): 228:1–228:25 (2025).
- [c24] Riccardo Benassi, Michele Luca Contalbo, Francesco Del Buono, Francesco Guerra, Giacomo Guiduzzi, Matteo Paganelli, Sara Pederzoli, Donato Tiano, Maurizio Vincini: **Explaining Entity Matching Models with CREW**. SEBD 2025: 570–579.
- [c23] Michele Luca Contalbo, Sara Pederzoli, Francesco Del Buono, Valeria Venturelli, Francesco Guerra, Matteo Paganelli: **GRI-QA: a Comprehensive Benchmark for Table Question Answering over Environmental Data**. Findings of ACL 2025: 15764–15779.
- [c22] Jan-Eric Hellenberg, Fabian Mahling, Lukas Laskowski, Felix Naumann, Matteo Paganelli, Fabian Panse: **PRISMA: A Privacy-Preserving Schema Matcher using Functional Dependencies**. EDBT 2025: 297–309.

2024

- [j6] Matteo Paganelli, Donato Tiano, Francesco Guerra: **A Multi-Facet Analysis of BERT-Based Entity Matching Models**. VLDB Journal 33(4): 1039–1064 (2024).
- [c21] Michele Luca Contalbo, Francesco Guerra, Matteo Paganelli: **Argument Relation Classification through Discourse Markers and Adversarial Training**. EMNLP 2024: 18949–18954.
- [c20] Matteo Paganelli, Donato Tiano, Francesco Del Buono, Andrea Baraldi, Riccardo Benassi, Giacomo Guiduzzi, Francesco Guerra: **How Transformers Are Revolutionizing Entity Matching**. SEBD 2024: 662–670.
- [c19] Riccardo Benassi, Francesco Guerra, Matteo Paganelli, Donato Tiano: **Explaining Entity Matching with Clusters of Words**. ICDE 2024: 2325–2337.
- [c18] Matteo Paganelli, Paolo Sottovia, Kwanghyun Park, Matteo Interlandi, Francesco Guerra: **Pushing ML Predictions into DBMSs (Extended Abstract)**. ICDE 2024: 5725–5726.

2023

- [j5] Matteo Paganelli, Paolo Sottovia, Kwanghyun Park, Matteo Interlandi, Francesco Guerra: **Pushing ML Predictions Into DBMSs**. IEEE Transactions on Knowledge and Data Engineering 35(10): 10295–10308 (2023).
- [c17] Andrea Baraldi, Francesco Del Buono, Francesco Guerra, Matteo Paganelli, Maurizio Vincini: **An Intrinsically Interpretable Entity Matching System**. EDBT 2023: 645–657.
- [c16] Marco Bergianti, Nicola Cioffo, Francesco Del Buono, Matteo Paganelli, Angelo Porrello: **Avoiding the Pitfalls on Stock Market: Challenges and Solutions in Developing Quantitative Strategies**. Ital-IA 2023: 489–494.
- [c15] Andrea Baraldi, Francesco Del Buono, Francesco Guerra, Giacomo Guiduzzi, Matteo Paganelli, Maurizio Vincini: **Interpretable Entity Matching with WYM**. SEBD 2023: 101–109.

2022

- [j4] Francesco Del Buono, Francesca Calabrese, Andrea Baraldi, Matteo Paganelli, Francesco Guerra: **Novelty Detection with Autoencoders for System Health Monitoring in Industrial Environments**. Applied Sciences 12(10): 4931 (2022).
- [j3] Matteo Paganelli, Francesco Del Buono, Andrea Baraldi, Francesco Guerra: **Analyzing How BERT Performs Entity Matching**. Proceedings of the VLDB Endowment 15(8): 1726–1738 (2022).
- [c14] Matteo Paganelli, Francesco Del Buono, Francesco Guerra, Nicola Ferro: **Evaluating the Integration of Datasets**. SAC 2022: 347–356.

- [c13] Andrea Baraldi, Francesco Del Buono, Matteo Paganelli, Francesco Guerra: **Landmark Explanation: a Tool for Entity Matching**. SEBD 2022: 69–76.
- [c12] Francesco Del Buono, Francesca Calabrese, Andrea Baraldi, Matteo Paganelli, Alberto Regattieri: **Data-Driven Predictive Maintenance in Evolving Environments: A Comparison Between Machine Learning and Deep Learning for Novelty Detection**. In *Sustainable Design and Manufacturing, Smart Innovation, Systems and Technologies* 262: 109–119 (2022).

2021

- [c11] Andrea Baraldi, Francesco Del Buono, Matteo Paganelli, Francesco Guerra: **Landmark Explanation: An Explainer for Entity Matching Models**. CIKM 2021: 4680–4684.
- [c10] Matteo Paganelli, Francesco Del Buono, Marco Pevarello, Francesco Guerra, Maurizio Vincini: **Automated Machine Learning for Entity Matching Tasks**. EDBT 2021: 325–330.
- [c9] Andrea Baraldi, Francesco Del Buono, Matteo Paganelli, Francesco Guerra: **Using Landmarks for Explaining Entity Matching Models**. EDBT 2021: 451–456.
- [c8] Matteo Paganelli, Paolo Sottovia, Antonio Maccioni, Matteo Interlandi, Francesco Guerra: **Using Descriptions for Explaining Entity Matches (Discussion Paper)**. SEBD 2021: 291–298.
- [c7] Francesco Del Buono, Matteo Paganelli, Paolo Sottovia, Matteo Interlandi, Francesco Guerra: **Transforming ML Predictive Pipelines into SQL with MASQ**. SIGMOD 2021: 2696–2700.

2020

- [j2] Matteo Paganelli, Paolo Sottovia, Antonio Maccioni, Matteo Interlandi, Francesco Guerra: **Explaining Data with Descriptions**. *Information Systems* 92: 101549 (2020).
- [c6] Matteo Paganelli, Francesco Del Buono, Francesco Guerra, Nicola Ferro: **Unsupervised Evaluation of Data Integration Processes**. iiWAS 2020: 77–81.

2019

- [j1] Matteo Paganelli, Domenico Beneventano, Francesco Guerra, Paolo Sottovia: **Parallelizing Computations of Full Disjunctions**. *Big Data Research* 17: 18–31 (2019).
- [c5] Paolo Sottovia, Matteo Paganelli, Francesco Guerra, Yannis Velegrakis: **Finding Synonymous Attributes in Evolving Wikipedia Infoboxes**. ADBIS 2019: 169–185.
- [c4] Francesco Guerra, Paolo Sottovia, Matteo Paganelli, Maurizio Vincini: **Big Data Integration of Heterogeneous Data Sources: The Re-Search Alps Case Study**. IEEE Big Data Congress 2019: 106–110.
- [c3] Matteo Paganelli, Paolo Sottovia, Antonio Maccioni, Matteo Interlandi, Francesco Guerra: **Understanding Data in the Blink of an Eye**. CIKM 2019: 2885–2888.
- [c2] Matteo Paganelli, Paolo Sottovia, Francesco Guerra, Yannis Velegrakis: **TuneR: Fine Tuning of Rule-Based Entity Matchers**. CIKM 2019: 2945–2948.

2017

- [c1] Francesco Guerra, Margherita Russo, Marco Fontana, Matteo Paganelli, François Bancilhon, Christian Frisch, Loïc Petit, Anna Giorgi, Emanuela Zilio: **The RE-SEARCH ALPS (Research Laboratories in the Alpine Area) Project**. HPCS 2017: 64–69.

Declaration

Last updated: August 2026