

CURRICULUM VITAE

Riccardo Cantini

Contents

1	General information	2
2	Positions Held	2
3	Education and Training	2
4	Research Activities	3
4.1	Deep Learning and Large Language Models	3
4.2	Distributed Big Social Data Analysis	4
5	List of Publications	5
6	Participation in Projects	7
7	Professional Services	7
8	Teaching Activities	8
9	Awards	8

1 General information

DATE OF BIRTH
30 December, 1993

PLACE OF BIRTH
Catanzaro (CZ), Italy

NATIONALITY
Italian

ADDRESS
Work: via Pietro Bucci 41C, 87036 Rende (CS), Italy
Home: via Teano n.13, 88100 Catanzaro (CZ), Italy

E-MAIL
riccardo.cantini@unical.it

2 Positions Held

- Since July 2023, *Researcher (RTDA) in Computer Engineering*, SSD ING-INF/05, at DIMES, University of Calabria, Rende (CS).
Scientific Supervisor: Prof. Domenico Talia.
- From March 2023 to July 2023, *Research Fellow* at DIMES, University of Calabria, Rende (CS), within the *ASPIDE* project.
Research Title: *Topic Detection and Tracking Techniques from Social Media Data*.
Scientific Supervisor: Prof. Fabrizio Marozzo.
- From April 2021 to July 2022, *Visiting Researcher* at the Department of Computer Science, Workflows and Distributed Computing Group, Barcelona Supercomputing Center (BSC), Spain.
Scientific Supervisor: Prof. Rosa M. Badia.
- From November 2019 to March 2023, *Ph.D. student in Information and Communication Technologies* at DIMES, University of Calabria, Rende (CS).
Scientific Supervisor: Prof. Paolo Trunfio, co-supervisor: Prof. Fabrizio Marozzo.
- From July 2019 to October 2019, *Research Fellow* at DIMES, University of Calabria, Rende (CS), within the *ASPIDE* project.
Research Title: *In-Memory Techniques for Efficient Execution of Data-Intensive Applications on Exascale Architectures*.
Scientific Supervisor: Prof. Paolo Trunfio.
- From July 2018 to June 2019, *Research Collaborator* at DIMES, University of Calabria, Rende (CS), within the *Smart Macingo* project.
Research Title: *Data Analysis Techniques for Price Estimation in Transportation Services*.
Scientific Supervisor: Prof. Paolo Trunfio.

3 Education and Training

- European Ph.D. in Information and Communication Technologies, DIMES, University of Calabria, Rende (CS).
Thesis Title: *Distributed big social data Analysis: Advanced Techniques and Execution Strategies*.
Supervisor: Prof. Paolo Trunfio.
Co-supervisor: Prof. Fabrizio Marozzo.
Evaluation: Excellent
Completion Date: 07/03/2023.

- Participation in the 1st International School on Internet of Things and Edge AI: Computing, Communications and Systems, Falerna (CZ).

Organizer: DIMES, University of Calabria, with the sponsorship of the IEEE Italy section and the Computer Engineering Group (GII).

Date: 08-12/09/2022.

- Master’s Degree in Computer Engineering, University of Calabria, Rende (CS).

Thesis Title: *Analysis of Political Polarization of Twitter Users through Neural Networks*.

Advisors: Prof. Paolo Trunfio, Prof. Fabrizio Marozzo.

Evaluation: 110/110 cum laude.

Completion Date: 12/04/2019.

- Bachelor’s Degree in Computer Engineering, University of Calabria, Rende (CS).

Thesis Title: *Using the ELKI Framework for the Discovery of Regions of Interest through Clustering Techniques*.

Advisors: Prof. Paolo Trunfio, Prof. Fabrizio Marozzo.

4 Research Activities

Riccardo Cantini’s research spans two distinct areas: *deep learning*, focusing on large language models (LLMs) and sustainable artificial intelligence, and *distributed big social data analysis*, targeting politically polarized data and the efficient execution of data-intensive applications in high-performance distributed environments. The following sections outline his main contributions to these fields.

4.1 Deep Learning and Large Language Models

Riccardo Cantini’s research in deep learning explores the potential of Transformer-based large language models (LLMs), such as BERT and GPT, showcasing their versatility across diverse domains. Sustainability is a central theme in this research area, emphasizing green awareness and promoting the efficient, fair, and trustworthy use of LLMs.

LLM Applications

Riccardo Cantini’s research in this field introduces several contributions, with a particular emphasis on social media contexts and real-world use cases. In the hashtag recommendation field, a novel BERT-based methodology was developed, namely *HASHET (Hashtag recommendation using Sentence-to-Hashtag Embedding Translation)* [9]. It employs dual latent spaces to embed the textual content of posts and their corresponding hashtags, learning to map semantic features of text to latent hashtag representations for a more accurate and scalable recommendation process. LLMs were also employed to tackle the spread of false information in online social media discourse [2], [16]. Specifically, a novel methodology was designed, termed *TM-FID (Topic-oriented Multimodal False Information Detection)*, which combines false information detection with neural topic modeling within a semi-supervised multimodal framework. By jointly leveraging textual and visual information in online news, TM-FID provides insights into how false information influences specific discussion topics, enabling a comprehensive and fine-grained understanding of its spread and impact on social media conversations. LLMs were also leveraged in automated reporting from social posts, with applications ranging from analyzing COVID-related discussions to enhancing emergency management during natural disasters [1], [15]. To ensure high-quality outputs, LLMs are enhanced with information across dimensions such as sentiment, emotion, and topic, extracted from user-generated content through fine-tuning and in-context learning approaches. The potential of LLMs has also been explored in the healthcare domain, investigating how explainable AI (XAI) models can present technical results in a human-readable format through seamless integration with generative AI, improving accessibility for healthcare professionals. This

approach was applied in two key areas: interpretable breast cancer grade prediction using genomic data and explainable depression detection from social media posts. In the former, explanations are achieved using interpretable-by-design machine learning models based on randomized trees. For the latter, explanations are generated using *BERT-XDD (BERT-based eXplainable Depression Detection)*, a novel self-explainable model that integrates BERTweet with a bi-directional Long Short-Term Memory (LSTM) network, producing classifications and explanations via masked attention.

Green-Aware and Sustainable AI

In this context, techniques for efficient fine-tuning are combined with curriculum learning, meta-learning, and knowledge distillation, enabling effective learning from limited data and scarce computing resources. Innovative methodologies have been proposed to enhance cross-architecture knowledge distillation from LLMs to more compute-efficient neural architectures. Notably, *DiXtill (XAI-driven Knowledge Distillation)* [3] introduces a novel approach to distilling explainable knowledge from a teacher LLM into a lightweight, self-explainable student model. This method combines local explanations with traditional prediction-based supervision, yielding higher accuracy and greater interpretability compared to standard distillation techniques. Furthermore, DiXtill achieves superior compression ratios and speedups compared to approaches like post-training quantization and attention head pruning, enabling efficient deployment on resource-constrained devices and advancing sustainable edge AI applications. Expanding upon online knowledge distillation, a parameter-efficient meta-distillation was also designed, combining the learning-to-teach paradigm with the efficiency of Low-Rank Adaptation (LoRA) within a second-order parameter-efficient learning framework.

Beyond computational efficiency, Cantini’s research also addresses social sustainability by examining biases and stereotypes in responses generated by LLMs across various scales [13]. Through adversarial analysis, hidden vulnerabilities in LLMs are probed, quantifying the effectiveness of different jailbreak attacks and exploring how model size influences filtering mechanisms and overall safety. By thoroughly characterizing LLM behavior under bias elicitation in terms of fairness and robustness, this analysis provides critical insights to promote the ethical design of AI systems. Similarly aimed at ensuring fairness in LLM output, retrieval-augmented generation (RAG) systems are explored, combining agentic designs with real-time retrieval and textual entailment to enhance the accuracy, fairness, and reliability of LLM responses. Finally, sustainability-related research encompasses the integration of green awareness into neural architecture search techniques and the development of interpretable energy estimation models for edge AI applications.

4.2 Distributed Big Social Data Analysis

Riccardo Cantini’s research in big social data analysis explores how detailed user information from social media can be leveraged to uncover users’ perceptions of real-world events, offering data-driven insights into socio-political phenomena. His work addresses critical issues including reliability, language barriers, and dynamicity, while also tackling the challenges related to resource-intensive computation.

Political Polarization in Big Social Data

Cantini’s research in this field is primarily centered on politically polarized big social data, mainly those generated during election campaigns and referenda. A primary focus is determining the political leanings of social media users by analyzing the intricate dynamics of the multifaceted online political discourse. Key contributions include *IOM-NN (Iterative Opinion Mining using Neural Networks)* [12], a semi-supervised methodology that employs feed-forward neural networks to estimate the political polarization of social users during political events. This tool offers an accurate, fast, and cost-effective alternative to traditional opinion polls, also capturing sentiments on sensitive topics that users might avoid addressing openly. This methodology was further developed to analyze the relationship between user polarization and sentiment toward political candidates, modeling political support across a broad spectrum of emotions. Topic modeling enhances this process, providing a richer representation of online conversations and enabling a comprehensive workflow for analyzing polarized data [7]. Cantini’s research also addressed challenges related to the influence of social bots on legitimate users and temporal dynamics in voting predictions. Specifically a

novel methodology was introduced, termed *TIMBRE* (*Time-aware Opinion Mining via Bot Removal*) [6], which estimates political polarization amidst temporal fluctuations and bot malicious activity, achieving a reliable estimate of user polarization. Furthermore, the integration of information diffusion and influence maximization with polarization analysis is explored through *WABC* (*Weighted Artificial Bee Colony*) [10], a bio-inspired algorithm designed to identify the key influencers in a politically polarized social network and uncover the primary information diffusion strategies employed by each faction during the political campaign.

Big Data Frameworks and ML-based Optimizations

This research addresses the computational challenges posed by the high volume and velocity of big social data by focusing on the efficient execution of data-intensive applications in high-performance distributed environments. A primary contribution is a detailed evaluation of scalable tools and paradigms [5], culminating in a published book on big data programming [20]. This book serves as a practical guide for developers, aiding in the selection of optimal solutions based on application requirements and available resources.

To optimize parallel and distributed applications, Cantini’s research introduces novel techniques for workflow task scheduling and data partitioning. For task scheduling, the *IWM* (*Intelligent In-memory Workflow Manager*) [11] was proposed, balancing task parallelism and memory usage to minimize data spilling and improve application throughput. By leveraging machine learning to predict task memory needs and execution times, IWM applies a heuristic bin-packing solution to generate effective schedules. Experiments conducted with Apache Spark demonstrate significant improvements in workflow performance compared to traditional approaches. In the area of data partitioning, a novel machine learning-based methodology was developed, named *BLEST-ML* (*BLOCK size ESTimation through Machine Learning*) [4]. This approach effectively estimates optimal data block sizes for hybrid partitioning, optimizing the execution of data-parallel applications on large-scale infrastructures with minimal resource and domain knowledge requirements. Evaluations using the dislib library of PyCOMPSs, conducted in diverse environments—including the MareNostrum 4 supercomputer at the Barcelona Supercomputing Center (BSC)—reveal substantial performance gains and reductions in execution time.

5 List of Publications

Journals

- [1] R. Cantini, C. Cosentino, F. Marozzo, D. Talia, and P. Trunfio, “Harnessing prompt-based large language models for disaster monitoring and automated reporting from social media feedback,” *Online Social Networks and Media*, vol. 45, p. 100 295, 2025.
- [2] R. Cantini, C. Cosentino, I. Kilanioti, F. Marozzo, and D. Talia, “Unmasking deception: A topic-oriented multimodal approach to uncover false information on social media,” *Machine Learning*, 2024, To appear.
- [3] R. Cantini, A. Orsino, and D. Talia, “Xai-driven knowledge distillation of large language models for efficient deployment on low-resource devices,” *Journal of Big Data*, vol. 11, no. 1, 2024. DOI: 10.1186/s40537-024-00928-3.
- [4] R. Cantini, F. Marozzo, A. Orsino, *et al.*, “Block size estimation for data partitioning in hpc applications using machine learning techniques,” *Journal of Big Data*, vol. 11, no. 1, 2024. DOI: <https://doi.org/10.1186/s40537-023-00862-w>.
- [5] L. Belcastro, R. Cantini, F. Marozzo, A. Orsino, D. Talia, and P. Trunfio, “Programming big data analysis: Principles and solutions,” *Journal of Big Data*, vol. 9, no. 1, 2022. DOI: 10.1186/s40537-021-00555-2.
- [6] R. Cantini, F. Marozzo, D. Talia, and P. Trunfio, “Analyzing political polarization on social media by deleting bot spamming,” *Big Data and Cognitive Computing*, vol. 6, no. 1, 2022. DOI: 10.3390/bdcc6010003.

- [7] L. Belcastro, F. Branda, R. Cantini, F. Marozzo, D. Talia, and P. Trunfio, “Analyzing voter behavior on social media during the 2020 us presidential election campaign,” *Social Network Analysis and Mining*, vol. 12, no. 1, 2022. DOI: 10.1007/s13278-022-00913-9.
- [8] L. Belcastro, R. Cantini, and F. Marozzo, “Knowledge discovery from large amounts of social media data,” *Applied Sciences*, vol. 12, no. 3, 2022. DOI: 10.3390/app12031209.
- [9] R. Cantini, F. Marozzo, G. Bruno, and P. Trunfio, “Learning sentence-to-hashtags semantic mapping for hashtag recommendation on microblogs,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 16, no. 2, 2021. DOI: 10.1145/3466876.
- [10] R. Cantini, F. Marozzo, S. Mazza, D. Talia, and P. Trunfio, “A weighted artificial bee colony algorithm for influence maximization,” *Online Social Networks and Media*, vol. 26, p. 100167, 2021. DOI: 10.1016/j.osnem.2021.100167.
- [11] R. Cantini, F. Marozzo, A. Orsino, D. Talia, and P. Trunfio, “Exploiting machine learning for improving in-memory execution of data-intensive workflows on parallel machines,” *Future Internet*, vol. 13, no. 5, 2021. DOI: 10.3390/fi13050121.
- [12] L. Belcastro, R. Cantini, F. Marozzo, D. Talia, and P. Trunfio, “Learning political polarization on social media using neural networks,” *IEEE Access*, vol. 8, pp. 47177–47187, 2020. DOI: 10.1109/ACCESS.2020.2978950.

Conferences

- [13] R. Cantini, G. Cosenza, A. Orsino, and D. Talia, “Are large language models really bias-free? assessing adversarial robustness to bias elicitation through jailbreak prompts,” in *International Conference on Discovery Science*, To appear, 2024.
- [14] P. Lindia, R. Cantini, F. Bettucci, L. Sartori, and P. Trunfio, “Enhancing the evaluation of fault detection models in smart agriculture using llm agents for rule-based anomaly generation,” in *1st Workshop on Green-Aware Artificial Intelligence, 23rd International Conference of the Italian Association for Artificial Intelligence*, To appear, 2024.
- [15] R. Cantini, C. Cosentino, and F. Marozzo, “Multi-dimensional classification on social media data for detailed reporting with large language models,” in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer, 2024, pp. 100–114.
- [16] R. Cantini, C. Cosentino, I. Kilanioti, F. Marozzo, and D. Talia, “Unmasking covid-19 false information on twitter: A topic-based approach with bert,” in *International Conference on Discovery Science*, Springer, 2023, pp. 126–140. DOI: https://doi.org/10.1007/978-3-031-45275-8_9.
- [17] R. Cantini and F. Marozzo, “Topic detection and tracking in social media platforms,” in *Pervasive Knowledge and Collective Intelligence on Web and Social Media*, C. Comito and D. Talia, Eds., Cham: Springer Nature Switzerland, 2023, pp. 41–56. DOI: 10.1007/978-3-031-31469-8_3.
- [18] R. Cantini, F. Marozzo, A. Orsino, M. Passarelli, and P. Trunfio, “A visual tool for reducing returns in e-commerce platforms,” in *2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI)*, IEEE, 2021, pp. 474–479. DOI: 10.1109/RTSI50628.2021.9597230.
- [19] L. Belcastro, R. Cantini, F. Marozzo, D. Talia, and P. Trunfio, “Discovering political polarization on social media: A case study,” in *2019 15th International Conference on Semantics, Knowledge and Grids (SKG)*, IEEE, 2019, pp. 182–189. DOI: 10.1109/SKG49510.2019.00038.

Books

- [20] D. Talia, P. Trunfio, F. Marozzo, L. Belcastro, R. Cantini, and A. Orsino, *Programming Big Data Applications: Scalable Tools and Frameworks for Your Needs*. World Scientific, 2023, ISBN: 978-1-80061-504-5. DOI: <https://doi.org/10.1142/q0444>.

Chapters

- [21] R. Cantini, F. Marozzo, and A. Orsino, “Deep learning meets smart agriculture: Using lstm networks to handle anomalous and missing sensor data in the compute continuum,” in *Device-Edge-Cloud Continuum - Paradigms, Architectures and Applications*, Springer Nature, 2023, pp. 141–153.
- [22] A. Orsino, R. Cantini, and F. Marozzo, “Evaluating the performance of a multimodal speaker tracking system at the edge-to-cloud continuum,” in *Device-Edge-Cloud Continuum - Paradigms, Architectures and Applications*, Springer Nature, 2023, pp. 155–166.

6 Participation in Projects

- *FAIR: Future Artificial Intelligence Research*
Research Objective: development of sustainable AI techniques focusing on interpretable energy estimation and the efficient, fair, and trustworthy use of Large Language Models.
- *eFlows4HPC: Enabling Dynamic and Intelligent Workflows in the Future EuroHPC Ecosystem*
Research Objective: use of machine learning techniques for optimizing data partitioning to support efficient execution of data-intensive workflows in HPC environments.
- *ASPIDE: exAScale ProgramIng models for extreme Data procEssing*
Research Objective: development of in-memory techniques for the efficient execution of data-intensive applications on Exascale architectures.
- *Smart Macingo*
Research Objective: definition and implementation of machine learning techniques for estimating the prices of transportation services.

7 Professional Services

Editorial Board

- Guest Editor for the special issue *Generative AI and Large Language Models* in the journal *Big Data and Cognitive Computing*.
- Review Editor of *Frontiers in Big Data*, Data Mining and Management section.
- Review Editor of *Frontiers in High-Performance Computing*.

Program Committee

- Program Chair of the *1st AIxIA Workshop on Green-Aware Artificial Intelligence (Green-Aware AI 2024)* workshop, co-located with the *23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024)*, Bolzano 25-28 November, 2024.
- PC member of *Advanced Data Mining and Applications (ADMA)* 2023, 2024 (Core rank: B).
- PC member of *Discovery Science (DS)* 2024 (Core rank: B).

Organization of Scientific Events

- Organizer of the *2nd FAIR Spoke workshop, Spoke 9 – Green-Aware AI*, Rende 24-25 March 2025.
- Organizer of the *1st AIxIA Workshop on Green-Aware Artificial Intelligence (Green-Aware AI 2024)* workshop, sponsored by the FAIR Foundation and co-located with the *23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024)*, Bolzano 25-28 November 2024.

Reviewer Activity

- International journals include: *ACM Computing Surveys*, *IEEE Transactions on Cloud Computing*, *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Big Data*, *Future Generation Computer Systems*, *Journal of Big Data*, *IEEE Access*, *Computer*, and *Social Network Analysis and Mining*.
- International conferences include: *International Conference on Advanced Data Mining and Applications* (ADMA), *International Conference on Learning Representations* (ICLR): 2024, *International European Conference on Parallel and Distributed Computing* (EuroPar), *IEEE International Conference on Machine Learning and Applications* (ICMLA), *IEEE International Conference on Big Data* (IEEE BigData), *International Conference in Parallel Processing* (ICPP), *International Conference on Discovery Science* (DS), and *IEEE International Conference on Bioinformatics and Biomedicine*.

8 Teaching Activities

- **High-Performance Computing**, Master's Degree in Computer Engineering, DIMES, University of Calabria, Rende (CS).
Topics Covered: main principles and practical aspects related to high-performance computing, with a focus on GPU computing in CUDA, and the MPI (Message Passing Interface) paradigm, encompassing both shared and distributed memory implementations.
A.Y. 2024-2025: 16 hours of lesson; A.Y. 2023-2024: 16 hours of lesson.
- **Operating Systems**, Bachelor's Degree in Computer Engineering, DIMES, University of Calabria, Rende (CS).
Topics Covered: main constructs for modeling and implementing multi-threaded applications, issues related to synchronization and access to shared variables, and concurrency mechanisms in Java, particularly Semaphores and Monitors.
A.Y. 2023-2024: 27 hours of lesson; A.Y. 2022-2023: 27 hours of lesson; A.Y. 2021-2022: 17 hours of lesson.
- **Distributed Systems and Cloud/Edge Computing for IoT**, Master's Degree in Computer Engineering for the Internet of Things, DIMES, University of Calabria, Rende (CS).
Topics Covered: fundamental concepts of Edge Computing in relation to Cloud architectures and the Internet of Things, with a focus on major tools and frameworks for modeling, simulating, and implementing large-scale interoperable IoT applications.
A.Y. 2021-2022: 23 hours of lesson; A.Y. 2020-2021: 23 hours of lesson; A.Y. 2019-2020: 23 hours of lesson.

Riccardo Cantini has also served as a **Thesis Advisor** for over 40 graduating students at DIMES, University of Calabria, overseeing both bachelor and master theses. The main topics of the supervised theses include machine and deep learning, natural language processing, large language models, sustainable artificial intelligence, and big data analysis.

9 Awards

- Ph.D. thesis *Distributed big social data Analysis: Advanced Techniques and Execution Strategies* recognized among the top 3 theses in *2024 Best Ph.D. and Master Thesis Awards in Big Data & Data Science*, organized by The CINI Data Science Lab.
- Journal article *Analyzing Political Polarization on Social Media by Deleting Bot Spamming* selected as the *issue cover* of *Big Data Cogn. Comput.*, Volume 6, Issue 1 (March 2022), and listed as an *Editor's Choice article* for its significance and impact in the field.

Declarations. I authorize the processing of my personal data included in the CV, in accordance with Article 13 of Legislative Decree no. 196 of June 30, 2003, “*Personal Data Protection Code*,” and Article 13 of GDPR 679/16, “*European Regulation on the Protection of Personal Data*”. I declare that the information provided in the CV is accurate and truthful. I declare to be aware of the implications of asserting the truthfulness of the above representation and to be informed of the criminal penalties under Article 76 of Legislative Decree no. 445 of December 28, 2000, “*Consolidated Text of Legislative and Regulatory Provisions Regarding Administrative Documentation*,” and in particular of what is provided for by Article 495 of the Criminal Code in case of false statements or false attestations. The above is presented in the form of a self-certification under Articles 19, 46, and 47 of Legislative Decree no. 445/2000.

Rende, 87036 Italy
December 29, 2024

Signature
RICCARDO CANTINI