Marta Marchiori Manerba

PhD in AI for Society

MIMOSA Research Fellow at Computer Science Department, University of Pisa

Research Associate KDD Laboratory at ISTI Institute, National Research Council in Pisa

Contacts

Birth: 11/3/96
Location: Based in Pisa & Turin (Italy)
Email: marta.marchiori@di.unipi.it
Website: `http://martamarchiori.github.io`
GitHub: `https://github.com/MartaMarchiori`
Scholar: Scholar Profile
ORCID: ORCID Profile
LinkedIn: LinkedIn Profile
Twitter: @Marta_Marchiori
Bluesky: @martamarchiori.bsky.social

Bio

Marta Marchiori Manerba (she/her) is a Graduate Student in Digital Humanities at the University of Pisa and is currently a Ph.D. student in AI. During her studies, she explored the relationship between technology and human rights. She works on Fairness and Explainability in Natural Language Processing, focusing on digital discrimination and algorithmic biases. She holds a Bachelor's Degree in Digital Humanities from the University of Pisa, during which she developed a strong interest in hate speech detection towards minorities in online discourse.

Research Interests

NLP; Responsible Language Technologies; Human-Centered AI; Explainability; Transparency; Fairness; Algorithmic Auditing; ML Evaluation; Data Awareness; Perspectivism; Intersectionality; Digital Discrimination; Abusive Language Detection

Education

NOVEMBER 2021 - July 2025
National PhD in AI for Society - *Computer Science Department, University of Pisa*
Supervisors: Riccardo Guidotti; Salvatore Ruggieri
Evaluation Panel: Su Lin Blodgett; Dino Pedreschi; Maurizio Tesconi
Doctoral fellowship on "Science and technology of interpretable machine learning and explanation of AI-assisted decision making" within the ERC Project XAI

FEBRUARY 2019 - JULY 2021
Master's Degree - *Digital Humanities: Language Technologies, University of Pisa*
Supervisor: Riccardo Guidotti
Mark: 110/110 with honors
Thesis project on Fairness Auditing in Abusive Language Detection Systems

OCTOBER 2015 - FEBRUARY 2019
Bachelor's Degree - *Digital Humanities, University of Pisa*
Supervisor: Maria Claudia Buzzi
Mark: 109/110
Thesis project carried out in the field of Assistive Technology for Education: development of a website for the research of tools and ICT resources to support cognitive training

Research & Academic Experience

NOVEMBER 2024 - OCTOBER 2025
Research Fellow (INFO-01/A) - *Computer Science Department, University of Pisa*
Research fellowship on "Design of Strategies for Evaluating, Refining and Mining Interpretable Models exploiting Sophisticated Algorithms considering Fairness, Privacy and Causality Aspects" under the FIS (Fondo Italiano per la Scienza) Project MIMOSA: Mining Interpretable Models explOiting Sophisticated Algorithms

5-6 JULY 2024
XAI-Hackathon - *Scuola Normale Superiore (Pisa)*
Winning team

JUNE-AUGUST 2024
Collaborator - *aequa-tech*
Together with the team of the start-up, that develops inclusive NLP technologies, I aimed to improve the transparency of the automated pipelines within the Debunker-Assistant, a web application that allows journalists and citizens to assess the trustworthiness of a newspaper article

APRIL-JULY 2023
Visiting Copenhagen Natural Language Understanding (CopeNLU) group at the Pioneer Centre for Artificial Intelligence - *Department of Computer Science, University of Copenhagen*
Supervisor: Isabelle Augenstein

OCTOBER - DECEMBER 2020
Traineeship - *Digital Humanities group, Fondazione Bruno Kessler*
Supervisor: Sara Tonelli
Fairness analysis of abusive language detection systems detecting unintended models biases with CheckList: creation of synthetic linguistic data to test a range of social prejudices (e.g., sexism, racism, and ableism)

MAY - SEPTEMBER 2018
Traineeship in Assistive Technology for Education - *Institute of Computer Science and Telematics, CNR*
Mapping and categorization of ICT resources to support cognitive training through the construction of an inventory of tools available on the internet; Implementation of a digital catalog to make the collection usable

Tutoring

12-13 JULY 2024
Data Analysis Tutor - *Traning Course of the FindHR (Fairness and Intersectional Non-Discrimination in Human Recommendation) Project*

SEPTEMBER 2023 - OCTOBER 2024
Co-Supervisor of a Master Thesis - *Data Science and Business Informatics, University of Pisa*
The contribution, titled "Fair and Explainable Clustering", was supervised alongside Riccardo Guidotti and Cristiano Landi

JANUARY-JULY 2022
Co-Supervisor of a Master Thesis - *Digital Humanities, University of Pisa*
The contribution, titled "Genetic Fairness-Enhancing Data Generation Framework", was supervised alongside Riccardo Guidotti and Martina Cinquini

OCTOBER - DECEMBER 2020
Academic Tutor of Programming and Theoretical Fundamentals - *Digital Humanities, University of Pisa*
Assistance during the laboratory: preparing exercises and assignments, helping students with coding during classes, clarifying basic constructs and concepts in Javascript

Publications

Conferences

[1] Eleonora Cappuccio et al. "An Interactive Interface for Feature Space Navigation". In: *HHAI 2024: Hybrid Human AI Systems for the Social Good*. IOS Press, 2024, pp. 73–83.

[2] Eleonora Cappuccio et al. "Beyond Headlines: A Corpus of Femicides News Coverage in Italian Newspapers". In: *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, Italy, December 4-6, 2024*. Ed. by Felice Dell'Orletta et al. Vol. 3878. CEUR Workshop Proceedings. CEUR-WS.org, 2024.

[3] Alessio Cascione et al. "Women's Professions and Targeted Misogyny Online". In: *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, Italy, December 4-6, 2024*. Ed. by Felice Dell'Orletta et al. Vol. 3878. CEUR Workshop Proceedings. CEUR-WS.org, 2024.

[4] Marta Marchiori Manerba and Riccardo Guidotti. “FairShades: Fairness Auditing via Explainability in Abusive Language Detection Systems”. In: *Third IEEE International Conference on Cognitive Machine Intelligence, CogMI 2021, Atlanta, GA, USA, December 13-15, 2021*. IEEE, 2021, pp. 34–43.

[5] Marta Marchiori Manerba and Riccardo Guidotti. “Investigating Debiasing Effects on Classification and Explainability”. In: *AIES ’22: AAAI/ACM Conference on AI, Ethics, and Society, Oxford, United Kingdom, May 19 - 21, 2021*. Ed. by Vincent Conitzer et al. ACM, 2022, pp. 468–478.

[6] Marta Marchiori Manerba et al. “Social Bias Probing: Fairness Benchmarking for Language Models”. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 14653–14671.

[7] Federico Mazzoni et al. “GenFair: A Genetic Fairness-Enhancing Data Generation Framework”. In: *Discovery Science - 26th International Conference, DS 2023, Porto, Portugal, October 9-11, 2023, Proceedings*. Ed. by Albert Bifet et al. Vol. 14276. Lecture Notes in Computer Science. Springer, 2023, pp. 356–371.

Journals

[1] Luca Nannini, Marta Marchiori Manerba, and Isacco Beretta. “Mapping the landscape of ethical considerations in explainable AI research”. In: *Ethics and Information Technology* 26.3 (2024), p. 44.

Workshops

[1] Marta Marchiori Manerba and Virginia Morini. “Exposing Racial Dialect Bias in Abusive Language Detection: Can Explainability Play a Role?” In: *PKDD/ECML Workshops (1)*. Vol. 1752. Communications in Computer and Information Science. Springer, 2022, pp. 483–497.

[2] Marta Marchiori Manerba and Sara Tonelli. “Fine-Grained Fairness Analysis of Abusive Language Detection Systems with CheckList”. In: *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*. Ed. by Aida Mostafazadeh Davani et al. Online: Association for Computational Linguistics, Aug. 2021, pp. 81–91.

[3] Marta Marchiori Manerba et al. “Bias Discovery within Human Raters: A Case Study of the Jigsaw Dataset”. In: *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP @LREC2022*. Ed. by Gavin Abercrombie et al. Marseille, France: European Language Resources Association, June 2022, pp. 26–31.

[4] Benedetta Muscato et al. “An Overview of Recent Approaches to Enable Diversity in Large Language Models through Aligning with Human Perspectives”. In: *Proceedings of the 3rd Workshop on Perspectivist Approaches to NLP (NLPerspectives) @ LREC-COLING 2024*. Ed. by Gavin Abercrombie et al. Torino, Italia: ELRA and ICCL, May 2024, pp. 49–55.

[5] Arianna Muti et al. “LeaningTower@LT-EDI-ACL2022: When Hope and Hate Collide”. In: *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Ed. by Bharathi Raja Chakravarthi et al. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 306–311.

[6] Mattia Setzu et al. “FairBelief - Assessing Harmful Beliefs in Language Models”. In: *Proceedings of the 4th Workshop on Trustworthy Natural Language Processing (TrustNLP 2024)*. Ed. by Kai-Wei Chang et al. Mexico City, Mexico: Association for Computational Linguistics, June 2024, pp. 27–39.

Extended Abstracts

[1] Isacco Beretta, Eleonora Cappuccio, and Marta Marchiori Manerba. “User-Driven Counterfactual Generator: A Human Centered Exploration”. In: *Joint Proceedings of the xAI-2023 Late-breaking Work, Demos and Doctoral Consortium co-located with the 1st World Conference on eXplainable Artificial Intelligence (xAI-2023), Lisbon, Portugal, July 26-28, 2023*. Ed. by Luca Longo. Vol. 3554. CEUR Workshop Proceedings. CEUR-WS.org, 2023, pp. 83–88.

[2] Marta Marchiori Manerba. “Eliciting Discrimination Risks in Algorithmic Systems: Taxonomies and Recommendations”. In: *Proceedings of the 3rd European Workshop on Algorithmic Fairness*. CEUR Workshop Proceedings. CEUR-WS.org, 2024.

[3] Marta Marchiori Manerba. “Fairness Auditing, Explanation and Debiasing in Linguistic Data and Language Models”. In: *xAI (Late-breaking Work, Demos, Doctoral Consortium)*. Vol. 3554. CEUR Workshop Proceedings. CEUR-WS.org, 2023, pp. 241–248.

Reviewing & Committees

Journals

- ACM Computing Surveys
- Machine Learning
- Information Processing and Management
- Language Resources and Evaluation

Conferences

- ACL Rolling Review (ARR)
- Conference on Empirical Methods in Natural Language Processing (EMNLP) 2023-2024
- Discovery Science (DS) 2024
- Italian Conference on Computational Linguistics (CLiC-it) 2024

Workshops

- Workshop on Perspectivist Approaches to NLP (NLPerspectives) 2022-2023
- ECML PKDD International Workshop on eXplainable Knowledge Discovery in Data Mining (XKDD) 2023, 2024
- Workshop on Language Technology for Equality, Diversity, Inclusion (LTEDI) EACL 2024
- Workshop on Algorithmic Fairness Through the Lens of Metrics and Evaluation (AFME) NeurIPS 2024
- ECML PKDD Workshop on Bias and Fairness in AI (BIAS) 2024
- European Workshop on Algorithmic Fairness (EWAF) 2024

Other Events

- International Symposium DataMod 2024 - From Data to Models and Back

Research Dissemination & Engagements

Events Organized

28 SEPTEMBER 2024
Laboratory organizer: "Tutelare i diritti nell'era dell'IA" (Protecting Rights in the Age of AI)
Event: Laboratory for high school students to collectively experience the potentials and limitations of chatbots, within the Scuola di Educazione Civica (Civic Education School) organized by Scuola Superiore Sant'Anna
in *Villa del Gombo*, *Parco di San Rossore*, *Pisa*, *Italy*

20-22 NOVEMBER 2023
Workshop organizer: "AI-GAP: Algorithmic Biases in Artificial Intelligence from Interdisciplinary Perspectives"
Event: Organized with funds won for PhD students' scientific initiatives, it addressed the social impact of AI technologies that reproduce human biases and inequities
at *Department of Humanities*, *University of L'Aquila*, *Italy*

16, 24 MARCH 2023
Laboratory organizer: "Does AI (NOT) Exist?"
Event: Laboratory to collectively experience the potentials and limitations of ChatGPT, organized within the CambiaMente Festival
in *Livorno*, *Italy*

3 FEBRUARY 2023
Panel organizer: "Picture a Scientist: A dialogue on stereotypes and gender gaps in STEM disciplines"
Event: Organized with funds won for PhD students' scientific initiatives, it aimed at creating awareness and fostering discussion about the gender gap in STEM
in *Pisa*, *Italy*

Talks & Panels

15 APRIL 2024
Round-table speaker: "Perchè Informatica a Pisa?" (Why Computer Science in Pisa?)
Event: "Incontra Informatica" (Meet Computer Science), orientation for high schools
at *Computer Science Department*, *University of Pisa*, *Italy*

24-26 NOVEMBER 2023
Speaker: "Privacy Network: Public Administration Audit and Advocacy for Digital Rights"
Event: Conference on AI for People: Democratizing AI
at *University of Bologna*, *Italy*

21 NOVEMBER 2023
Round-table speaker: "Dialogue on human and algorithmic stereotypes: is artificial intelligence neutral?"
Event: Part of the AI-GAP workshop and organized along with the association Fuori Genere, it was focused on the topic of Bias in AI, with particular emphasis on linguistics bias
in *L'Aquila*, *Italy*

27 MARCH 2023
Speaker: “Mind the GAP: Gender Gap in STEM Disciplines”
Event: High school civic education lesson
at *ITCG Cerboni*, *Portoferraio*, *Italy*

23 MAY 2022
Speaker: “Ethics in AI Systems: Biases in ML Models”
Event: High school civic education lesson
at *IIS Cobianchi*, *Verbania*, *Italy*

2 MARCH 2022
Speaker: “Ethics in AI Systems and the Timnit Gebru case”
Event: Orientation for high schools
at *Computer Science Department*, *University of Pisa*, *Italy*

Other

20-22 JANUARY 2021
Assistant moderator: “DH for Society – Equality, participation, rights and values in the digital age”
Event: Conference of the Italian Association for Digital Humanities and Digital Culture (AIUCD 2021)
held virtually

27 SEPTEMBER 2024
Assistant: Bright Night 2024
Event: “Càndidati”
in *Pisa*, *Italy*

Volunteering

14-16 OCTOBER 2024
Volunteer: “International Conference on Discovery Science” (DS 2024)
in *Pisa*, *Italy*

4-6 DECEMBER 2024
Volunteer: “Italian Conference on Computational Linguistics” (CLiC-it 2024)
in *Pisa*, *Italy*

Additional Training

12-15 MARCH 2024
Winter School on Foundation Models (Amsterdam, Netherlands)
by *ELLIS*

26-28 JULY 2023
Doctoral Consortium (Lisbon, Portugal)
within the *1st International Conference on eXplainable Artificial Intelligence*

4-8 JULY 2022
AI & Society 2022 Summer School (Pisa, Italy)
by *National Ph.D. in Artificial Intelligence for Society*

MAY-JUNE 2022
PhD course on Interdisciplinary Approaches for Bias Elicitation (Pisa, Italy)
by *NoBIAS: AI without Bias Marie Skłodowska-Curie Innovative Training Network*

JANUARY 2022
Winter Course on Deep Learning for Natural Language Processing (held virtually)
by *Ixa and HiTZ*

21-23 OCTOBER 2021
D as Digital Festival (Rovereto, Italy)
by *Informatici Senza Frontiere (Computer scientists without frontiers)*
Winner of a scholarship

14-18 JUNE 2021
Nordic Probabilistic AI School (held virtually)
by *Norwegian Open AI Lab and Norwegian University of Science and Technology*

JANUARY 2021
Elements of AI and **Ethics of AI** (online courses)
by *University of Helsinki*

DECEMBER 2020
Language, gender identity and Italian (online course)
by *Eduopen Network with Ca' Foscari University of Venice*

27-29 SEPTEMBER 2020
Scientific Research Festival: The future of science and humans in the age of augmented intelligence (Trieste, Italy)
by *Trieste Next*
Participation at the Academy Program

1-7 SEPTEMBER 2019
Summer School on Information Science (Osijek, Croatia)
by *European Information Science Education*
Winner of a scholarship

Digital Advocacy

2022 - IN PROGRESS
Research & Advocacy Officer - *Privacy Network*
Non-profit association that promotes a culture of privacy and responsible use of technology and that advocates for digital rights

2019 - 2023
Vice president and board member (2022, 2023) - *KRINO*
Cultural association that organizes panels, workshops, and events within the field of Digital Humanities, to integrate and contaminate sciences and arts

Volunteering

SEPTEMBER 2017 - OCTOBER 2022
Scout educator - *Association of Italian Guides and Scouts (AGESCI)*
Association contributing to the education and personal development of youth according to scouting principles and methods

Dichiarazione Sostitutiva di Certificazione
Dichiarazione Sostitutiva dell'Atto di Notorietà
Artt. 46 e 47 del D.P.R. 445/2000

La sottoscritta Marta Marchiori Manerba, codice fiscale MRCMRT96C51E897J, nata a Mantova (MN) il 11/03/1996, residente a Livorno (LI) in Via Eugenia 9

Visto il D.P.R. 28 Dic. 2000, n. 445 concernente T.U. delle disposizioni legislative e regolamentari in materia di documentazione amministrativa e successive modifiche ed integrazioni;

Vista la Legge 12 Nov. 2011, n. 183 ed in particolare l'art. 15 concernente le nuove disposizioni in materia di certificati e dichiarazioni sostitutive;

Consapevole che, ai sensi dell'art. 76 del D.P.R. 445/2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono punite ai sensi del Codice penale e delle leggi speciali vigenti in materia;

Dichiara sotto la propria responsabilità che quanto indicato nel presente curriculum vitæ et studiorum, comprensivo delle informazioni sulla produzione scientifica, corrisponde a verità.

La sottoscritta è consapevole della responsabilità penale prevista dall'art. 76 del D.P.R. 445/2000, per le ipotesi di falsità in atti e dichiarazioni mendaci ivi indicate.

Ai sensi e per gli effetti dell'Art.13 del decreto legislativo 30 giugno 2003, n.196, la sottoscritta autorizza al trattamento dei dati personali.