# Marta Marchiori Manerba *(she/her)*

## 2° Year PhD Student in AI for Society - ERC's Project XAI

### About

Marta Marchiori Manerba (she/her) is a Graduate Student in Digital Humanities at the University of Pisa and 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.

### Contacts

11/3/96 • Based in Pisa, Italy •
marta.marchiori@phd.unipi.it •
http://martamarchiori.github.io •
Skype • LinkedIn • @Marta_Marchiori

### Research Interests

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

---

### Education

NOVEMBER 2021 - ONGOING
**National PhD in AI for Society** - Computer Science Department, University of Pisa
Supervisors: Riccardo Guidotti; Salvatore Ruggieri. 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: Dr. 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: Dr. 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 and Academic Experience

JANUARY-JULY 2022
**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

2022
**Reviewer** - LREC Workshop on Perspectivist Approaches to NLP (NLPerspectives)

20-22 JANUARY 2021
**Assistant moderator** - Association for Digital Humanities and Digital Culture Conference (AIUCD2021)

OCTOBER - DECEMBER 2020

**Traineeship** - Digital Humanities group, Fondazione Bruno Kessler
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)

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 concept in Javascript

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 catalogue to make the collection usable

---

## Publications

### *Conference Publications*

*In-person presentation*
Marta Marchiori Manerba and Riccardo Guidotti. "Investigating Debiasing Effects on Classification and Explainability". AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES), 2022

*Online presentation, Best Paper Award*
Marta Marchiori Manerba and Riccardo Guidotti. "FairShades: Fairness Auditing via Explainability in Abusive Language Detection Systems". IEEE Conference on Cognitive Machine Intelligence (CogMI), 2021

### *Workshop Publications*

*In-person presentation*
Marta Marchiori Manerba and Virginia Morini. "Exposing Racial Dialect Bias in Abusive Language Detection: Can Explainability Play a Role?". ECML PKDD International Workshop on eXplainable Knowledge Discovery in Data Mining (XKDD), 2022

*In-person presentation*
Marta Marchiori Manerba, Riccardo Guidotti, Lucia Passaro, Salvatore Ruggieri. "Bias Discovery Within Human Raters: A Case Study of the Jigsaw Dataset". LREC Workshop on Perspectivist Approaches to NLP (NLPerspectives), 2022

Arianna Muti, Marta Marchiori Manerba, Katerina Korre, and Alberto Barrón-Cedeño. "LeaningTower@LT-EDI-ACL2022: When Hope and Hate Collide". ACL Workshop on Language Technology for Equality, Diversity and Inclusion (LT-EDI), 2022

*Online presentation*
Marta Marchiori Manerba and Sara Tonelli. "Fine-Grained Fairness Analysis of Abusive Language Detection Systems with CheckList". Workshop on Online Abuse and Harms (WOAH!), 2021

### *Currently under review*

Federico Mazzoni, Marta Marchiori Manerba, Martina Cinquini, Riccardo Guidotti, and Salvatore Ruggieri. "FairGen: A Genetic Fairness-Enhancing Data Generation Framework". Intelligent Data Analysis (IDA), 2023

---

## Research Dissemination: Engagements & Talks

**2022-2023**
Organisation of a panel discussion titled "Picture a Scientist: A dialogue on stereotypes and gender gaps in STEM disciplines" - Computer Science Department, University of Pisa
Won funds for PhD students' scientific initiatives. The event aims at creating awareness and fostering discussion about the gender gap in STEM

**MAY 2022**
High school civic education lesson tilted "Ethics in AI Systems: Biases in ML Models" - IIS Cobianchi in Verbania, Italy

**MARCH 2022**
University orientation event for high schools: talk titled "Ethics in AI Systems and the Timnit Gebru case" - Computer Science Department, University of Pisa

**MARCH 2022**
Video for university choice orientation on Digital Humanities and the impacts of AI - "A Choice for Life"

---

## Digital Advocacy

**NOVEMBER 2022 - IN PROGRESS**
Team member - Privacy Network
Non-profit association that promotes a culture of privacy and responsible use of technology and that advocates for digital rights

**NOVEMBER 2019 - IN PROGRESS**
Vice president and board member (2022) - KRINO
Cultural association that organizes panels, workshops and events within the field of Digital Humanities, with the aim of integrating and contaminating sciences and arts
-   Active participation in projects of critical divulgation of topics such as: computational thinking, datafication, digital culture, demystification of AI systems, digital art, live coding
-   Collaborative and collective writing and editing of posts, articles, newsletters
-   Internal training sessions

---

## Memberships

**2020 - IN PROGRESS**
Algorithmic Justice League

**FEBRUARY 2021 - DECEMBER 2021**
Associazione Italiana di Linguistica Computazionale (Italian Association of Computational Linguistics)

**JANUARY 2021 - DECEMBER 2021**
Associazione per l'Informatica Umanistica e la Cultura Digitale (Association for Digital Humanities and Digital Culture)

---

## 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
-   Organisational and coordination skills; ability to work in teams and on projects
-   Communication and relational skills developed as educator; aptitude for listening and confrontation