

Proceedings of "Ethics and Trust in Human-AI Collaboration: Socio-Technical Approaches", an IJCAI 2023 workshop

M. Bergamaschi Ganapini¹, A. Loreggia², N. Mattei³, F. Rossi⁴, B. Srivastava⁵ and K. B. Venable⁶

¹Union College - USA

²University of Brescia - Italy

³Tulane University - USA

⁴IBM Research - USA

⁵University South Carolina - USA

⁶University West Florida, IHMC - USA

Abstract

This volume contains the papers presented at the Ethics and Trust in Human-AI Collaboration: Socio-Technical Approaches, that was held on August 21, 2023, MACAO, China.

These papers were selected by an international program committee among all those submitted to the symposium through an open call.

1. Aim of the workshop

It is increasingly acknowledged that AI needs to be used to augment human intelligence, rather than replacing it. It is reasonable and useful to automate some tasks, but most of the tasks will be tackled by combining the complementary capabilities of humans and machines.

This is the case for "classical" AI models like classifiers and predictors, that are aimed to support human decision making, especially in high-risk domains. Even the most recent AI advances, like those in generative AI, exploit the power of language or other kinds of content to allow AI systems to better interact with humans and to support their creativity.

However, for this collaboration to work well, special attention needs to be put in designing hybrid systems in a way that trust and ethics issues are addressed satisfactorily. Without trust, which requires assets such as misinformation detection, explainability, transparency, and fairness, humans will not fully exploit the available AI capabilities. We also need to address ethics issues, such as the possible deskilling and displacement of human decision makers and the


Ethics and Trust in Human-AI Collaboration: Socio-Technical Approaches, August 21, 2023, MACAO, China

✉ bergamam@union.edu (M. B. Ganapini); andrea.loreggia@unibs.it (A. Loreggia); andrea.loreggia@unibs.it (N. Mattei); Francesca.Rossi2@ibm.com (F. Rossi); BIPLAV.S@sc.edu (B. Srivastava); bvenable@uwf.edu (K. B. Venable)

🆔 0000-0002-9846-0157 (A. Loreggia); 0000-0002-3569-4335 (N. Mattei); 0000-0001-8898-219X (F. Rossi); 0000-0002-7292-3838 (B. Srivastava); 0000-0002-1092-9759 (K. B. Venable)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

risk of value misalignment. Trust and ethics need to be a central and integral aim of the design, development, and use of human-AI collaboration systems, just as decision quality. We need to have better decisions than humans or machines alone, while achieving trust and resolving ethical issues.

To achieve this, we can and should exploit knowledge of how humans make decisions and interact with others (humans or artificial agents). Thus cognitive theories or knowledge from other cognitive sciences are essential to achieve these goals.

This workshop aims to connect three main areas of study: Human-AI collaborative environments, Ethics and Trust, and Cognitive theories of the human mind. We envision an audience of scholars from at least these three disciplines, that can interact and exchange ideas and solutions.

The symposium website can be found at this URL: <https://sites.google.com/view/ethaics-2023/>

2. Program Committee

The program committee included the following researchers, who reviewed papers and made the final decisions about acceptance for presentation and inclusion in the proceedings. We received a total of 8 submissions, of which 6 have been accepted.

- Claudia Passos Ferreira (New York University)
- Cristina Cornelio (Samsung Research)
- Lirong Xia (RPI)
- Michele Loi (Politecnico di Milano)
- Sujoy Sikdar (Binghamton University)
- Toby Walsh (UNSW Sydney)
- Umberto Grandi (IRIT, Université Toulouse Capitole)

3. Organizing Committee

The workshop was organized by the following researchers, with the support of the IJCAI office.

- Marianna Bergamaschi Ganapini (Union College)
- Andrea Loreggia (University of Brescia)
- Nicholas Mattei (Tulane University)
- Francesca Rossi (IBM Research)
- Biplav Srivastava (AI Institute)
- Brent Venable (University of South Florida and IHMC)