MULTITTRUST - Multidisciplinary Perspectives on Human-AI Team Trust

Centeio Jorge, Carolina; Ulfert-Blank, Anna Sophie

Publication date
2023

Document Version
Final published version

Published in
CEUR Workshop Proceedings

MULTITTRUST - Multidisciplinary Perspectives on Human-AI Team Trust

Carolina Centeio Jorge[1], Anna-Sophie Ulfert-Blank[2]

[1]Delft University of Technology, Delft, Netherlands
[2]Eindhoven University of Technology, Eindhoven, Netherlands

Abstract

This preface summarises the first Workshop on Multidisciplinary Perspectives on Human-AI Team Trust (MULTITTRUST 2023), co-located with 2nd International Conference on Hybrid Human-Artificial Intelligence (HHAI 2023), held on June 26th 2023 in Munich, Germany.

1. Introduction

This workshop appears from the need to create a multidisciplinary research community focused on studying the different perspectives and layers of trust dynamics in human-AI teams. Human-AI teamwork is no longer a topic of the future. With the increasing prominence of these teams in diverse industries, several challenges arise that need to be addressed carefully. The study of trust has a longstanding tradition across disciplines (e.g., human-computer interaction or psychology). Yet, understanding how trust is defined and how it functions in Human-AI teams remains a challenge. Psychological literature suggests that within human teams, team members rely on trust to make decisions and to be willing to rely on their team. Besides that, the multi-agent systems (MAS) community has been adopting trust mechanisms to support the decision-making of the agents regarding their peers. Finally, in the last couple of years, researchers have been focusing on how humans trust AI and how AI can be trustworthy. But when we think of a team composed of both humans and AI, with recurrent (or not) interactions, complex dynamics, and diverse team compositions, how do these theories and findings all come together? Currently, we are missing approaches that integrate prior literature on trust in teams across disciplines (esp. Psychology and Computer Science). In particular, when looking at dyadic or team-level trust relationships in such teams, we also need to look at how an AI should trust a human teammate and how trust can be defined. Furthermore, human trust in the AI team member and trust by AI agents in human team members will change over time and also affect each other. In this workshop, we wanted to motivate the conversation across the different fields and domains.

HHAI-WS 2023: Workshops at the Second International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 26—27, 2023, Munich, Germany
*Corresponding author.
C.Jorge@tudelft.nl (C. Centeio Jorge); A.S.Ulfert.Blank@tue.nl (A. Ulfert-Blank)
https://research.tudelft.nl/en/persons/carolina-centeio-jorge (C. Centeio Jorge)
0000-0002-6937-5359 (C. Centeio Jorge); 0000-0001-6293-4173 (A. Ulfert-Blank)

Together, we may shape the road to better defining what trust in human-AI teams entails and resolve future questions.

This workshop called for contribution and/or participation from several disciplines, including Psychology, Sociology, Cognitive Science, Computer Science, Artificial Intelligence, Robotics, Human-Computer Interaction, Design, and Philosophy. Topics related to this workshop include:

- Measures of team trust in human-AI teams.
- Human's trust and trustworthiness in human-AI teams.
- Dynamics of trust between human and AI in teamwork.
- Hybrid techniques (knowledge-driven + data-driven) to assess trust and trustworthiness in human-AI teams.
- Machine learning techniques to detect trust and trustworthiness in human-AI teams and teammates.
- Evaluation methods for trust and trustworthiness models in human-AI teams.
- Experimental settings for trust dynamics in human-AI teams.
- Design of systems that take into account trust dynamics in human-AI teams.

2. Organization

2.1. Workshop Chairs

- Carolina Centeio Jorge, Delft University of Technology, NL
- Anna-Sophie Ulfert-Blank, Eindhoven University of Technology, Eindhoven, NL

2.2. Programme Committee

- Filipa Correia, ITI-LARSYS, PT
- Cristiano Castelfranchi, ISTC-CNR, IT
- Alessandro Sapienza, ISTC-CNR, IT
- Michelle Zhao, Carnegie Mellon University, US
- Rino Falcone, ISTC-CNR, IT
- Catholijn Jonker, Delft University of Technology, NL
- Siddharth Mehrotra, Delft University of Technology, NL
- Beau Schelble, Clemson University, US
- Filippo Cantucci, ISTC-CNR, IT
- Mengyao Li, University of Wisconsin-Madison, US
- X. Jessie Yang, University of Michigan, US
- Connor Esterwood, University of Michigan, US
- Samuele Vinanzi, Sheffield Hallam University, UK
- Alan R. Wagner, Penn State University, US
- Ewart de Visser, USAFA, US
- Glenda Hannibal, Ulm University, DE
- Hebert Azevedo-Sá, Military Institute of Engineering, BR
- Eleni Georganta, University of Amsterdam, NL
- Ruben Verhagen, Delft University of Technology, NL

3. Programme

In this workshop, we wanted to provide the space required for building a multidisciplinary community. With that in mind, we had a combination of presentations and interaction moments, both for networking and discussing the related topics. Around twenty people attended the workshop and several expressed their appreciation for such a format to discuss important topics that need input from several disciplines, as they recognized the challenges that persist regarding the study of trust in human-AI teams.

The day started out with a networking activity where participants could play *human bingo*. In this activity, each participant had a grid with random facts about people, for example "Someone wearing glasses", and they could go around trying to find the name of another participant with such a trait. The room quickly became lively, and the ice was broken. We believe this was important to make participants more comfortable to ask questions and open discussion.

The rest of the day consisted of two keynote talks and five sessions of paper presentations. Our two keynote speakers were Prof. Lionel P. Robert from University of Michigan, and Dr. Myrthe Tielman from Delft University of Technology. Finally, each paper session consisted of two short lightning talks (seven minutes each, without Q&A) followed by sixteen minutes of discussion about the overarching topic of the session. These discussion moments at the end of each session were crucial for the engagement of the audience and for allowing a deeper connection and argumentation.

3.1. Keynote Talks

- *The Problematic Problems of Human Trust in Robots: Is Trusting a Robot More like a Teammate or a Tool and should we really care?* by Lionel P. Robert Jr. from University of Michigan.
  Abstract: As robotics advances and permeates various aspects of our social and work lives, the question of how humans we view and ultimately trust robots has become increasingly pertinent. Do humans view them as mere machines, automated tools designed to serve their needs or do they embrace a more empathetic approach, viewing and trusting them as actual teammates (i.e. humans)? On the one hand, proponents of robots as possible humans argue that computers are social actors (CASA) and that humans mindlessly interact with computers in much the same way they do humans. This view is often used to justify the employment of human-to-human theories and their corresponding measures to understand human-robot interactions. On the other hand, advocates of mechanization contend that humans do not view robots as humans but instead as automated tools. This view discourages using human-to-human theories and their corresponding measures to understand human-robot interactions. They advocate for more human-to-automation theories and measures of constructs like trust. In this thought-provoking presentation, I will explore the arguments supporting both perspectives and consider the potential consequences of each approach. Ultimately, this presentation aims to provide a balanced understanding of the complexities involved to encourage a nuanced dialogue on the subject.
- *Let's talk about trust* by Myrthe L. Tielman from Delft University of Technology.

Abstract: Trust is a hot topic. It's something very important to humans, it's important to teams, and it's important for AI. So many people are looking into trust, and as human-AI team researchers it seems something we should care a lot about. But what do we actually mean when we talk about trust? There's a lot of different perspectives and definitions. Should we care about that, or try to come to an agreement? In this talk, I argue that meaning is more important than agreement when it comes to words. But meaning is crucial, as through looking at the different meanings of trust, we also might gain new perspectives on how to achieve it.

3.2. Paper Sessions (Lightning Talks)

The Programme Committee (PC) received 12 submissions of short abstract papers (one column, 2 to 4 pages, excluding references). Each paper was carefully reviewed by three reviewers based on its relevance to the workshop and writing. It was not required to include novel elements in the paper but rather to summarise the authors' line of research and their contribution to the community. In the end, ten papers were accepted for presentation. They were divided into five topics (which formed paper sessions): *Perception of AI teammate's trustworthiness*, *Improving AI teammate's trustworthiness*, *Calibrating Human-AI trust in teams*, *Decision-making in Human-AI teams*, and *Human-AI Team Trust*.

- **Perception of AI teammate's trustworthiness**
  - *The Trustworthiness Assessment Model – A Micro and Macro Level Perspective* by Nadine Schlicker and Markus Langer.
  - *AI-Enabled Decision Support Systems: Tool or Teammate?* by Myke C. Cohen and Michelle Mancenido.
- **Improving AI teammate's trustworthiness**
  - *Communicating AI intentions to boost Human AI cooperation* by Bruno Berberian, Marin Le Guillou and Marine Pagliari.
  - *The Effects of Social Intelligence on Trust in Human-AI Teams* by Morgan Bailey, Benjamin Gancz and Frank Pollick
- **Calibrating Human-AI trust in teams**
  - *Investigating Human-Robot Overtrust During Crises* by Colin Holbrook, Daniel Holman, Alan Wagner, Tyler Marghetis, Gale Lucas, Brett Sheeran, Vidullan Surendran, Jared Armagost, Savanna Spazak, Kevin Andor and Yinxuan Yin. *Unfortunately, none of the authors could present, given a last minute paperwork impediment.*
  - *Mutually Adaptive Trust Calibration in Human-AI Teams* by Ewart de Visser, Ali Momen, James Walliser, Spencer Kohn, Tyler Shaw and Chad Tossell.
- **Decision-making in Human-AI teams**
  - *Causing Intended Effects in Collaborative Decision-Making* by André Meyer-Vitali and Wico Mulder.
  - *Artificial Trust for Decision-Making in Human-AI Teamwork: Steps and Challenges* by Carolina Centeio Jorge, Catholijn M. Jonker and Myrthe L. Tielman.

- **Human-AI Team Trust**
  - *Trust Dispersion and Effective Human-AI Team Collaboration: The Role of Psychological Safety* by Tilman Nols, Anna-Sophie Ulfert-Blank and Avi Parush.
  - *Piecing Together the Puzzle: Understanding Trust in Human-AI Teams* by Anna-Sophie Ulfert-Blank, Eleni Georganta, Myrthe L. Tielman and Tal Oron-Gilad.

Acknowledgments

This research was supported by Delft AI Initiative, the SIOP Visionary Grant, and by EU Horizon 2020 research and innovation programme under GA Numbers 952215 (TAILOR) and 820437 (Humane AI Net), and supported by the National Science Foundation (NWO) under Grant Number 024.004.022 (Hybrid Intelligence). The support is gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting organisations.

Finally, the organisers and authors would like to thank HHAI 2023 team, in particular the workshop chairs, for organising and providing the infrastructure that made MULTITTRUST 2023 possible.