Preface to MULTITTRUST 3.0 - Multidisciplinary Perspectives on Human-AI Team Trust*

Myrthe L. Tielman[1,*,†], Andre Meyer-Vitali[2,†], Morgan Bailey[3,†] and Francesco Frattolillo[4,†]

[1]*Delft University of Technology, Delft, the Netherlands*
[2]*DFKI, Saarbrucken, Germany*
[3]*University of Glasgow, Glasgow, UK*
[4]*Sapienza University of Rome, Rome, Italy*

Abstract

The third edition of the workshop on Multidisciplinary perspectives on Human-AI Team Trust (MULTITTRUST) was held at the International conference on Hybrid Human-Artificial Intelligence at Malmö on June 11, 2024. This workshop brought together interdisciplinary researchers working on the topic of trust in Human-AI teams to present their work, and discuss the most prominant challenges in this field.

Keywords

HHAI, Trust, Human-AI teamwork, Human-AI interaction

1. Introduction

This workshop originates from the need to create a multidisciplinary research community of people who study the different perspectives and layers of trust dynamics in teams consisting of both humans and AI agents. Two successful previous editions of this workshop ran in 2023, and with MULTITTRUST 3.0 we are aiming to take this momentum to continue building this community. Human-agent teamwork is no longer a topic of the future. With the increasing prominence of human-agent interaction in hybrid teams in diverse industries, several challenges arise that need to be addressed carefully. One of these challenges is understanding how trust is defined and how it functions in Human-agent teams. 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 decision-making of the agents regarding their peers and for delegating tasks to agents. Finally, in the last couple of years, researchers have been focusing on how humans trust AI agents and how such systems can be trustworthy. But when we think of a team composed of both humans and agents, with recurrent (or not) interactions, how do these all come together? Currently, we are missing approaches that integrate prior literature on trust in

HHAI-WS 2024: Workshops at the Third International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 10—14, 2024, Malmö, Sweden

*Corresponding author.
†These authors contributed equally.
m.l.tielman@tudelft.nl (M. L. Tielman); Andre.Meyer-Vitali@dfki.de (A. Meyer-Vitali); m.bailey.1@research.gla.ac.uk (M. Bailey); frattolillo@diag.uniroma1.it (F. Frattolillo)

teams in these different disciplines (esp. Psychology and Computer Science). In particular, when looking at dyadic or team-level trust relationships in such a team, we also need to look at how an AI should trust a human teammate. In this context, trust, or rather the factors that influence it, must be formally defined so that the AI can evaluate them, rather than using questionnaires at the end of a task, as is usually assessed in psychology. Furthermore, the trust of the human in the artificial team member and vice-versa will change over time and also affect each other. In this workshop, we want to motivate the conversation across the different fields and domains. Together, we may shape the road to answer these questions and more.

2. Output of the workshop

In this workshop, aimed to provide the space for building a multidisciplinary community. With this aim, our workshop included some elements of traditional workshops, but also some of a more reflective event where discussions and interactions are the main elements in our format.

We received 5 submissions of **extended abstracts**, all of which were accepted and presented at the workshop in a lightning talk. The goal of these papers was to serve as a base for introducing the participant's to each-others work and expertise. These submission showed an interesting mix of perspectives on the topic, including experimental designs, presentations of early stage work, studies on human trust and studies on artificial trust. Additionally, we received 2 submissions of **progress papers**, which were both accepted, but only one which was presented at the workshop in a short talk. The goal of these submissions was to allow attendees to share new results and insights, as well as discuss work in progress.

Additionally, Professor Bertram F. Malle presented a **keynote** on his work on Multi-Dimensional Trust: History, Evidence, and Applications to Individuals and Systems. There is much talk about trust in AI and robots; but it often seems unclear what trust really is, and how we should measure it. This talk briefly reviewed some strands of work that to the development of a multi-dimensional model and measure of trust that integrates numerous previous approaches. The talk then showed that ordinary people have a multi-dimensional conception of trust, whose primary dimensions are competence and reliability (together forming Performance trust), as well as ethical integrity, sincerity, and benevolence (together forming Moral trust). Finally, the talk introduced the MDMT, an intuitive measure of multi-dimensional trust, and presented evidence for its validity and usefulness in studying trust in numerous targets, such as people, machines, animals, and institutions.

In the afternoon, three **discussion groups** were formed around three main topics: components of meaningful studies; concepts and definitions; and the role of agents and artificial trust. Within these groups, structured discussions were had to identify the main current challenges in the field from different perspectives. These discussions attempted to start a more structured analysis of the research questions that are prevalent in the field. The workshop finished with the groups sharing the main output of their discussions with the larger group. The organizing team aims to try to consolidate this output further to strengthen the community.

3. Program Committee

This section contains a list of our program committee, we thank them for their contributions to the reviewing process.

- Connor Esterwood, University of Michigan, US - Confirmed
- Siddharth Mehrotra, Delft University of Technology, NL - Confirmed
- Michelle Zhao, Carnegie Mellon University, US - Confirmed
- Mengyao Li, University of Wisconsin-Madison, US - Confirmed
- Alan R. Wagner, Pennsylvania State University, US - Confirmed
- Beau Schelble, Clemson University, US - Confirmed
- Frank Pollick, University of Glasgow, UK - Confirmed
- Piercosma Bisconti, DEXAI – Artificial Ethics, IT - Confirmed
- Angelo Cangelosi, University of Manchester, UK - Confirmed
- Antonio Chella, University of Palermo, IT
- Antonella Marchetti, Universita' Cattolica del Sacro Cuore di Milano, IT
- Sivia Rossi, Federico II University of Naples, IT
- Alessandra Rossi, Federico II University of Naples, IT