Modeling Trust in Human Robot Interactions(HRI)

Csenge Hubay[1], Musfira Jilani[2], Fernando Perez Tellez[2]

1.Budapest University of Technology, Budapest, Hungary | 2. Technological University Dublin, Dublin, Ireland

Abstract. Trust is often regarded as a driving force of socialisation. Even though trust has been studied interdisciplinarily, there is no consensus on its real-time measurement in the context of HRI. This research aims to implement a system using motion capture and video data to assess the trust of people in robots real-time. With this, robots could give accurate feedback to the reactions of people and facilitate cooperation.

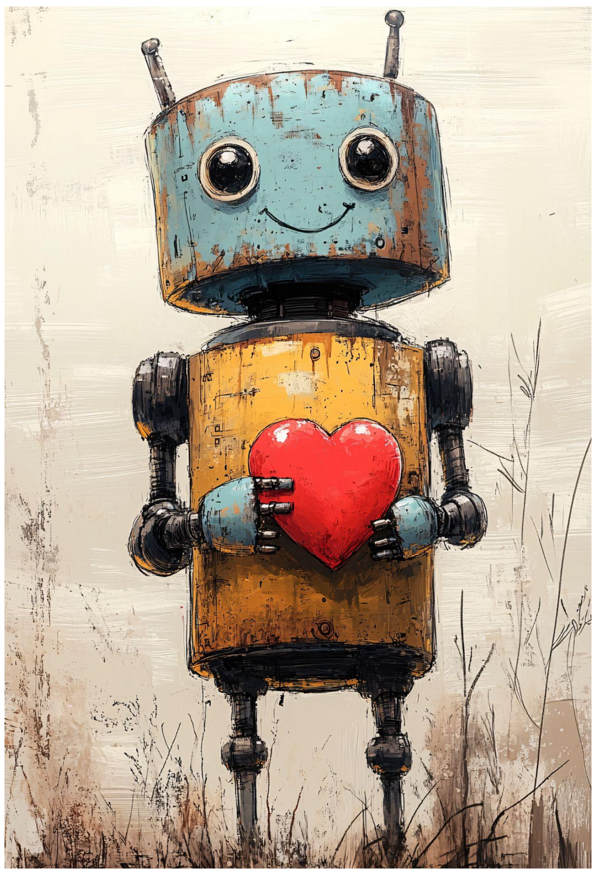

Figure 1, source: Pixabay

INTRODUCTION

As cooperation with robots becomes more widespread, it is important to develop ideal conditions for HRI. Trust is a key factor in this coexistence, as it influences safety performance and the human agents' willingness to collaborate [1].

Trust has been studied in various disciplines and is often seen as an essential connective part of society. Even though scholars agree on its importance, there is no consensual definition for the term. For most, it indicates expectations towards the other, not behaviour patterns. Cooperative behaviour is promoted and induced by these expectations. To achieve confidence in machines the first step is to provide an accurate metric to measure the trust. Self-reports give post-hoc insights and do not aid real-time assessment. Other disadvantages of this method are that there is no uniform questionnaire to assess trust and that self-reports often turn out to be subjective.
The main questions of this research are whether **trust in HRI be reliably assessed through real-time analysis of human and robot behavioral data**, and **what factors best predict trust levels across different interaction contexts**? Our hypothesis is that **behavioral data**, such as movement capture **can aid the identification of the key elements of trust**.

LITERATURE REVIEW

- **Trust assessment in other studies**
  - ***Self-reports*** yield post-hoc insights based on the feelings of the subjects [2]
  - ***Physiological signals*** **(EEG, GSR)** used in trust assessment studies achieved ***70%+ accuracy*** [3,4].
- **Movement data approach:**
  - Motion capture systems or video-based analysis
  - XSENS sensors applied in motion classification [8,9].

DATA TO USE

Two eligible datasets

- **DrapeBot benchmark dataset [5]**
  - XSENS motion capture data
  - Industrial environment
  - Trust in robots scored by test subjects
- **SSUP-HRI dataset [6]**
  - 360°camera footage
  - Urban environment
  - No trust scoring
  - Further preprocessing needed (eg. with PoseFlow [7])

RESEARCH METHODOLOGY

The research planned consists of the following steps:

1. **Identify Behavioral Trust Indicators**: Analyze movement and social signaling data to identify key behavioral markers linked to trust in HRI settings.
2. **Develop a Trust Prediction Model**: Utilize machine learning to build predictive models based on the behavioral and motion tracking data.
3. **Evaluate Model Performance Across Contexts**: Test the trust prediction model on diverse datasets (e.g., SSUP-HRI and Drapebot) to ensure robustness across public and industrial settings.
4. **Explore Real-time Trust Modulation**: Investigate how trust levels fluctuate dynamically within sessions, potentially adjusting robotic behaviors in real-time to foster trust.

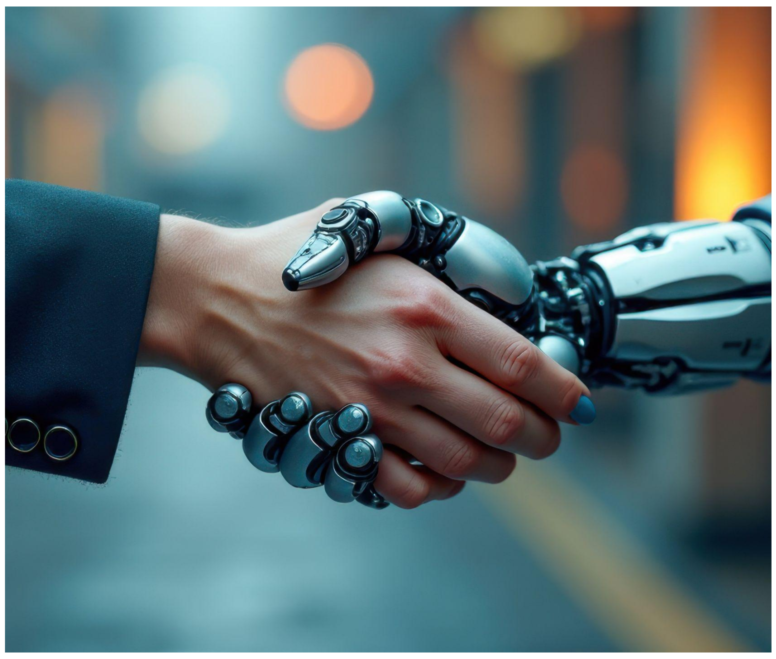

Figure 2, Source: Pixabay

PRELIMINARY CONSIDERATIONS

With the implementation of a system, capable of **real-time trust measurement**. research on robots giving **feedback** to mistrust detected would potentially be possible. This would **facilitate cooperation** and cohabitation with robots and aid their social acceptance.

CONCLUSIONS

The aim of this research is to create a system capable of measuring the trust of humans working with robots and other AI-driven agents based on their motion.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This poster was created as part of the Blended Intensive Programme organized under the Erasmus + Programme of the European Union

References.

[1] Khavas, Z. R., Ahmadzadeh, S. R., & Robinette, P. (2020, November). Modeling trust in human-robot interaction: A survey. In International conference on social robotics (pp. 529-541). Cham: Springer International Publishing.
[2] Bauer, P. C., & Freitag, M. (2018). Measuring trust. The Oxford handbook of social and political trust, 15.
[3] Xu, C., Zhang, C., Zhou, Y., Wang, Z., Lu, P., & He, B. (2024). Trust Recognition in Human-Robot Cooperation Using EEG. arXiv preprint arXiv:2403.05225.
[4] Akash, K., Hu, W. L., Jain, N., & Reid, T. (2018). A classification model for sensing human trust in machines using EEG and GSR. ACM Transactions on Interactive Intelligent Systems (TiiS), 8(4), 1-20.
[5] Rehm, M., Hald, K., & Pontikis, I. (2024, March). Benchmark Movement Data Set for Trust Assessment in Human Robot Collaboration. In Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (pp. 934-938).
[6] Bu, F., & Ju, W. (2024). SSUP-HRI: Social Signaling in Urban Public Human-Robot Interaction dataset. arXiv preprint arXiv:2403.10994
[7] Xiu, Y., Li, J., Wang, H., Fang, Y., & Lu, C. (2018). Pose Flow: Efficient online pose tracking. arXiv preprint arXiv:1802.00977.
[8] Shin, H. C., Soo, H. K., Lee, D. W., & Ki, S. Y. (2023, October). Lower body action classification using unlabeled predicted motion. In 2023 14th International Conference on Information and Communication Technology Convergence (ICTC) (pp. 1858-1860). IEEE.
[9] Mathur, D., & Bhatia, D. (2022). Gait classification of stroke survivors-An analytical study. Journal of Interdisciplinary Mathematics, 25(1), 163-181.