Workshop on the Design of Responsible Hybrid Intelligence

Ludi van Leeuwen, Cor Steging and Bart Verheij

Bernoulli Institute of Mathematics, Computer Science and Artificial Intelligence, University of Groningen

Abstract

We summarize the first Workshop on the Design of Responsible Hybrid Intelligence (RHI2023), co-located with the 2st International Conference on Hybrid Human-Artificial Intelligence (HHAI 2022), held on June 27, 2022 in Munich, Germany.

Keywords

Responsible AI, Hybrid Intelligence

Applications of hybrid intelligence systems in which humans and machines collaborate must behave responsibly in order to promote synergy and prevent unwanted or even harmful effects. For instance, the agents in a hybrid intelligence system (both human and automated) should communicate in a correct and reliable way, they should be able to provide their reasons for having a belief or opinion, and they should be able to explain their actions in terms of the values they apply.

In practice, it proves to be a hard and complex problem to design responsible hybrid intelligence systems. For instance, both humans and machines can make mistakes and be unreliable, have unjustified beliefs and positions, and can act without considering their values or even go against them.

A current and telling example of a hybrid intelligence system is ChatGPT in conversation with a human user, which has provided a new level of natural hybrid interaction about a wide range of topics. Many have experienced that the conversation contains correct and reliable elements, but also mistakes and unreliable behavior. Also the reasons provided by ChatGPT can be helpful and convincing, but also irrelevant or vacuous. Furthermore ChatGPT refers to following a value system when avoiding harmful or sensitive topics, but also can act against the values expressed in an erratic way.

Against this background, we organized the first Workshop on the Design of Responsible Hybrid Intelligence (RHI2023), held on June 27, 2022 in Munich, Germany. The event was co-located with the 2st International Conference on Hybrid Human-Artificial Intelligence (HHAI 2022).

HHAI-WS 2023: Workshops at the Second International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 26—27, 2023, Munich, Germany

The authors contributed equally.

l.s.van.leeuwen@rug.nl (L. van Leeuwen); c.c.steging@rug.nl (C. Steging); bart.verheij@rug.nl (B. Verheij)

https://sites.google.com/rug.nl/ludivanleeuwen/about (L. van Leeuwen); https://steging.nl/ (C. Steging); https://www.ai.rug.nl/~verheij/ (B. Verheij)

0000-0003-3165-4376 (L. van Leeuwen); 0000-0001-6887-1687 (C. Steging); 0000-0001-8927-8751 (B. Verheij)

© 2023 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)

CEUR-WS.org/Vol-3456/preface2.pdf

Topics and issues on designing responsible hybrid intelligence we aimed to address included, but were not limited to, the following:

- Methods and tools for designing responsible hybrid intelligence systems
- Theoretical foundations for the design of responsible hybrid intelligence systems
- Experiments with and innovative applications of responsible hybrid intelligence systems
- Evaluation methods for implemented responsible hybrid intelligence systems
- The role of various AI approaches in the design of responsible hybrid intelligence systems (data, knowledge, reasoning, argumentation, language, ...)

The aim of the workshop was to collect research and research ideas aimed at designing responsible hybrid intelligence systems. Hence we aimed to collect contributions and discussions that bridge the technological and ethical side of responsible hybrid intelligence design. For all contributors (whether technological or ethical), we asked to make explicit how and to what extent the research or research idea discussed aims to contribute to responsible hybrid intelligence design. We invited submissions of all levels of maturity (early stage, mid stage, completed). A selection was made on the basis of overall quality, relevance and diversity.

The workshop included two keynote lectures, one by Stefan Schlobach (Free University Amsterdam) and one by Lameck Mbangula Amugongo (Technische Universität München). A diverse set of papers were presented. The program was aimed at stimulating discussions between the participants, an interdisciplinary group of researchers, ranging in focus from technical implementation to societal implications of AI. The workshop concluded with a plenary discussion led by Virginia Dignum (Umeå University, Delft University of Technology) on how we could and should design Responsible Hybrid Intelligence, acknowledging the challenges and encouraging design.

The Program Committee (PC) received a total of 6 submissions. Following a single-blind reviewing process, each paper was peer-reviewed by at least two PC members. The committee decided to accept 5 short papers, containing original work.

The specifics of the program were as follows.

Keynotes

- Stefan Schlobach - *Responsible Hybrid Intelligence Systems—a Knowledge Representation Perspective*
- Lameck Mbangula Amugongo - *Designing and Implementing Responsible AI*

Paper presentations

- Annet Onnes, Silja Renooij, Mehdi Dastani - *Bayesian Network Conflict Detection for Normative Monitoring of Black-Box Systems*
- Jonne Maas - *Not a black box, but an empty one: Accounting for power in AI systems*
- Marco Zullich and Giovanni Santacatterina - *Assessing Fairness in Open-Source Face Mask Detection Algorithms*
- Tijn van der Zant - *Trias Intelligentia*
- Virginia Dignum, Petter Ericson - *Towards a feminist, relational, conception of Artificial Intelligence*

Plenary Discussion

- Virginia Dignum - Moderator

Organization

Workshop Chairs

- Ludi van Leeuwen, University of Groningen
- Cor Steging, University of Groningen
- Bart Verheij, University of Groningen

Program Comittee

- Annet Onnes, Utrecht University
- Petter Ericsson, Umeå University
- Jonne Maas, Delft University
- Wijnand van Woerkom, Utrecht University
- Rineke Verbrugge, University of Groningen
- Henry Prakken, Utrecht University & University of Groningen
- Michael Dale, Eindhoven University of Technology

Acknowledgments

The organisation of the workshop was supported by the Hybrid Intelligence Center, a 10-year programme funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research, https://hybrid-intelligence-centre.nl. The organizers would like to thank the HHAI 2023 workshop chairs and organization for providing an excellent framework for RHI2023.