Can You Trust Your Robotic Assistant?

Farshid Amirabdollahian[1], Kerstin Dautenhahn[1], Clare Dixon[2], Kerstin Eder[3,4], Michael Fisher[2], Kheng L. Koay[1], Evgeni Magid[3,4], Anthony Pipe[3,5], Maha Salem[1], Joe Saunders[1], Matt Webster[2]

[1] Adaptive Systems Research Group, U. Hertfordshire [adapsys.feis.herts.ac.uk]
[2] Centre for Autonomous Systems Technology, U. Liverpool [www.liv.ac.uk/cast]
[3] Bristol Robotics Laboratory [www.brl.ac.uk/vv]
[4] Computer Science Dept., U. Bristol [www.bristol.ac.uk]
[5] Engineering, Design and Mathematics Dept., U. West of England [www.uwe.ac.uk]

Abstract. Robotic assistants are being developed to assist with a range of tasks at work and home. Besides designing and developing such robotic assistants, a key issue that needs to be addressed is showing that they are both safe and trustworthy. We discuss our approach to this using formal verification, simulation-based testing and formative user evaluation.

Keywords: Safety; trust; verification; service robotics; robotic assistants.

1 Introduction

Robotic Assistants are now being designed to help us at work and at home, e.g., in our everyday activities or in health-care scenarios. A robotic assistant can do a variety of things, from simply fetching your glasses or helping you put together flat-pack furniture, to running a bath, and even analysing your health needs. A wide range of robotic assistants is currently under development within academia and industry, from surgery and rehabilitation robots to flexible manufacturing robots. These all indicate a rise in development and use of robotic technologies.

At the present time, the major challenge no longer lies in producing such robotic helpers, but in demonstrating that they are *safe* and *trustworthy*. The Trustworthy Robotic Assistants project[1] aims to address exactly this. Specifically, we propose a holistic approach to developing and providing a unified safety framework for human–robot interaction (HRI). It combines our previous experiences with *formal verification* (e.g., [4]), *simulation-based testing* (e.g. [3]) and *formative user evaluation* (e.g. [1, 2]). Based on this combination of techniques, we aim to tackle the holistic analysis of safety and trustworthiness in HRI.

2 Research Approach

In practice, no single verification/validation technique is adequate to cover all safety aspects of an HRI scenario. Therefore the analysis of trustworthiness in robotic assistants requires a combination of different techniques.

Formal verification can exhaustively analyse all of the robot's possible choices, but uses a vastly simplified environmental model. Simulation-based testing of

[1] http://www.robosafe.org/

human–robot interactions can be carried out in a fast, directed way and involves a much more realistic environmental model, but is essentially selective and does not take into account true human interaction. Formative user evaluation provides exactly this validation, constructing a comprehensive analysis from the human participant's point of view. Though non-trivial to achieve, this combined approach will be very powerful. Not only will analysis from one technique stimulate new explorations for the others, but each distinct technique actually remedies some of the deficiencies of the others. Therefore, this combination provides a new, strong, comprehensive, end-to-end verification and validation method for assessing safety in human-robot interactions.

We start with the user requirements, which are transformed into a system specification that formally describes the HRI framework and characterizes all components of the framework: the assumptions about the human behaviour, the robot and the environment, the HRI protocols, and the expected outcomes of interactions between these components. The key components of the specification, presented as explicit requirements, include the safety of the human and the environment. The system is designed according to the specification and will be implemented with real robots, based on *ROS*[2] platforms. Formal verification will verify that the design is correct with respect to the specification, and simulation-based methods will verify the system design against the implementation. To strengthen our approach, simulation-based methods will be applied to verify the implementation against the specification. Finally, the end-user evaluation will validate our implementation with regard to the initial user requirements.

We will be developing a unified safety framework for HRI. This framework can be used to generate evidence of the correctness of robotic assistants, allowing them to gain the trust of their users. The safety and trustworthiness components provided by the Trustworthy Robotic Assistants project will inform future safety standards and will eventually contribute to professional safety assessment.

Acknowledgements. The authors were partially supported by the Trustworthy Robotic Assistants[1] project funded by EPSRC grants EP/K006193/1, EP/K006320/1, EP/K006509/1 and EP/K006223/1.

References

1. K. Dautenhahn, Methodology and Themes of Human-Robot Interaction: A Growing Research Field, *Int. J. of Advanced Robotic Systems*, vol. 4(1): 103–108, 2007.
2. E. C. Grigore, K. Eder, A. Pipe, C. Melhuish and U. Leonards, Joint Action Understanding improves Robot-to-Human Object Handover, *IEEE IROS*, 2013.
3. E. Magid, T. Tsubouchi, E. Koyanagi, Y. Yoshida and T. Tadokoro, Controlled Balance Losing in Random Step Environment for Path Planning of a Teleoperated Crawler Type Vehicle, *J. of Field Robotics*, vol. 28(6): 932–949, 2011.
4. R. Stocker, L. A. Dennis, C. Dixon and M. Fisher, Verification of Brahms Human–Robot Teamwork Models, *Proc. 13th European Conf. on Logics in Artificial Intelligence*, LNCS, vol. 7519: 385–397, Springer 2012.

[2] Robot Operating System, http://www.ros.org/