Journal of Responsible Technology 14 (2023) 100059

Contents lists available at ScienceDirect

Journal of Responsible Technology

journal homepage: www.sciencedirect.com/journal/journal-of-responsible-technology

Reflections on Responsible Research and Innovation (RRI) for Trustworthy Autonomous Systems (TAS): A message from Journal of Responsible Technology Special Issue's editors

Responsible Research and Innovation (RRI) consists of a set of well-grounded principles and processes that stimulate openness, reflection and stakeholder engagement in research projects, programmes and institutions. RRI involves considering the potential ethical, social, and environmental implications of new technologies and taking steps to address any potential negative impacts identified. The development of autonomous systems, driven by ubiquitous computing, the growth in the digital economy, advances in robotics, and rapid developments in artificial intelligence, should include RRI as a pre-requisite and starting point.

For this special issue, we specifically invited researchers to submit short reflective pieces that discussed experiences of applying RRI or empirical studies concerned with applying RRI to the field of developing Trustworthy Autonomous Systems (TAS). We sought papers that critically considered the barriers and facilitators of 'doing' RRI, and how action plans were deployed both successfully and unsuccessfully. We asked for lessons learnt for future work and for defining RRI best practices. Submitting authors were expressly instructed *not* to include hypothetical discussions of how RRI should be framed or considered, unless this was accompanied by a real-world example.

We recognise that reflective papers in the style we sought are rare in peer-reviewed journals; we therefore provided our reviewers the following questions to consider for assessing the quality and suitability of the manuscript reviewed:

- When putting RRI into practice, how do the authors interpret their experiences?
- Do the authors describe clearly what happened and reflect on an actual experience?
- Do the authors describe what was positive or negative about the experience?
- Do they suggest solutions, new ideas, or recommendations as a result of this, and are these sensible and well thought out?
- Is the relevance of RRI to TAS clear?

This special issue is our 'message in a bottle' from 12 contributions in which researchers reflected on their RRI journeys. We urge readers to find it. Open it. Reflect and act on its content. We wish to thank all the invited persons for their generosity in first saying yes and then following through across several months in preparing these specific reflexive papers.

The first paper by Joseph Lindley and colleagues titled 'Towards a Master Narrative for Trust in Autonomous Systems: Trust as a Distributed Concern' explores the role of Trust in RRI. Trust is a central element for autonomous systems linked to algorithmic explainability, accountability and transparency; systems' verification, validation, and reliability; governance and regulation. In order to synthesise the multitude of perspectives which exist on Trust, the authors apply qualitative methods to create a 'Master Narrative' to unify and guide thoughts, beliefs, values and behaviours. This Master Narrative is defined as 'Trust as a Distributed Concern' and operationalised when applying RRI in the context of TAS. This approach does not answer the question 'Is this system Trustworthy'? instead, it creates new ways to interrogate and reflect about TAS challenges including frameworks to structure complex information that allow researchers to explore new context dependent narratives on TAS.

Stevienna de Saille and her team discuss playful approaches to engage with RRI within their TAS funded project 'Imagining Robotic Care: Identifying conflict and confluence in stakeholder imaginaries of autonomous care systems'. In their article 'Using LEGO® SERIOUS® Play with stakeholders for RRI' the authors note that anticipatory and reflexive practices should begin at the problem-definition stage, involving a wide range of stakeholders. However, early pre-award engagement activities are difficult to conduct without pre-existing funding and resources. Without stakeholder opinions to influence the early stages of a project, researchers can unintentionally introduce biases and assumptions. LEGO® SERIOUS® Play is an inexpensive, accessible means of exploring divergent needs, assumptions, capacities and constraints. It is playful methodology to bring lesser-heard voices into the processes of innovation and focus on what values should govern designing practices.

A third paper exploring multidisciplinary and disruptive approaches to RRI is from Pauline Leonard and Chira Tochia, titled 'From episteme to techne: Crafting responsible innovation in trustworthy autonomous systems research practice'. The authors highlight important aspects of research that too often are ignored, for example, the complex and multidimensional issues of power and emotion which are especially important when researching trust and trustworthiness. Through their TAS project "Trustworthy Human-Robot Teams' there is an acknowledgement that trust is not just a rational calculation, but a process that is context-specific and difficult to quantify. There is an urge to become more aware and critical of conceptual frameworks, positions, biases, political affiliations, expectations and justifications. All these impact on the framing of research questions, decisions taken on research methods

Abbreviations: AI, Artificial Inteligence; IT, Information Technology; RRI, Responsible Research and Innovation; TAS, Trustworthy Autonomous Systems.

https://doi.org/10.1016/j.jrt.2023.100059

Available online 24 January 2023

and participants, the collection of data, the interpretation of findings, and the presentation of knowledge. By revisiting the AREA framework, the authors skilfully invite the reader to embrace RRI practices by learning from feminist research, and the interplay of politic, power and emotion within the research process. There are conceptual tools adopted by feminist/qualitative researchers that could be useful for enhancing responsible research and anticipating the multiplicity of social issues that influence research practices.

Tara Roberson and colleagues explore the value of RRI when developing and deploying TAS in defence. 'A method for ethical AI in defence: A case study on developing trustworthy autonomous systems' focuses on the barriers that industry faces when engaging in RRI - time and money - but also the benefits: mitigating ethical risks, reducing the adverse humanitarian effects of warfare, and demonstrating responsible behaviours to their stakeholders. Within the Athena AI' project (i.e., a tool to identify objects and people who cannot be targeted in the battlefield) as a case study, researchers reflect on three RRI dimensions; embedding ethics-by-designed principles, applying governance RRI frameworks and the power of stakeholder engagement. The authors highlight the importance of *reflexibility* to ensure responsible decision-making, agile governance frameworks that promote *responsiveness* to meet legal and ethical requirements and *inclusion* to assess levels of trust. This is an excellent example of an industry-lead approach for RRI in action outside academic innovation.

'Responsible research and innovation in practice: Driving both the 'How' and the 'What' to research' is an excellent applied paper from Chen and colleagues that demonstrates that RRI can be an active catalyst for shaping research. Rather than being perceived as a 'speed bump' constraining the research process, RRI can support researchers to see alternative possibilities regarding the nature and scope of their initiatives. RRI can indeed guide how research is being conducted as a *research safeguard*, as well as acting as a *research driver* to articulate new research ideas and topics to be studied. The authors suggest whole-team participation in collective reflective exercises to boost a culture of responsibility around TAS.

Carolyn Ten Holter and colleagues provide a case study to explore responsibility around data. In their article 'Responsible Innovation; responsible data. A case study in autonomous driving', authors explore RRI challenges including politics, power imbalances, values, and conceptual questions of what 'responsibility' means. Specifically, they focus on stakeholders and how broad engagement can resolve some of the issues that impact on autonomous vehicles data (what data to collect, who can access, etc.). The authors recommend projects to be assessed on the RI approach and operationalisation, consider sustainability and carbon impact, and finally engagement with policymakers. An important point is to embrace RRI as a iterative and flexible process that permits incorporating adjustments in response to findings.

'Involving psychological therapy stakeholders in responsible research to develop an automated feedback tool: Learnings from the ExTRAPPOLATE project' is an exemplar for embedding RRI within the development of a feedback tool for therapists called Auto-CICS. Jackob A. Andrews and colleagues bring together an inclusive and representative group of patients and practitioners in a series of online workshops designed to identify concerns and recommendations relevant for the design of Auto-CICS. The authors do a magnificent job illustrating the challenges and difficulties encountered during these workshops and how criticisms were skilfully transformed into learning outcomes, insights and corrective actions to feed and refine the development of the Auto-CICS tool.

Helen Smith and colleagues reflect on how they added RRI elements to their research practices by expanding EDI (Equality, Diversity & Inclusion) principles on participant recruitment and other research activities. A simple change on their participants' demographic questionnaire meant being more inclusive and less alienating to those from ethnic and gender minorities, that often do not fit into a pre-defined category being forced to tick 'other', instead of self-describing their gender or ethnicity. The authors identify 'inclusivity' as a force for change for the future development of novel TAS.

The paper titled 'Reflections on RRI in "TAS for Health at Home"' by Nils Jaeger and colleagues focuses on the experiences and value that emerged from a group of multidisciplinary researchers applying the AREA framework to investigate a smart mirror system for healthcare. The authors clearly map how RRI tools (I.e., Moral IT cards, engagement activities such as workshops with multidisciplinary experts and Patient and Public Involvement groups) support the identification of challenges and solutions useful to inform technology development and deployment. Through the lenses of RRI, this paper highlights the *home* as a unique environment for considering known data privacy issues, anticipating purposes and unintended uses, and interphase design challenges that are idiosyncratic to specific medical conditions.

In their article, 'Supporting responsible research and innovation within a university-based digital research programme: Reflections from the hoRRIzon project', Virginia Portillo and colleagues, reflect on the challenges that researchers encounter when putting RRI into practice. These include the time and timing required to engage meaningfully with RRI practices, frequent confusion with research ethics and integrity policies, and the importance of institutional support such as training and award schemes. Understanding the value of RRI and appreciating its relevancy, are important aspects to be considered when promoting RRI practices among the TAS research community and beyond.

Richard Waterstone and colleagues present an excellent study of how telepresence robots can contribute to open research. In 'Robot telepresence as a practical tool for responsible and open research in trustworthy autonomous systems', the authors propose that an Open Laboratory approach can increase the transparency of scientific research by making it easier for the general public to access it. For example, cameras on robots can provide an insightful view to the wider public about lab activities in real time, while contributing to a better understanding about how these technologies work. Open Science can improve technical awareness and research literacy by making research processes more visible, understandable and trustworthy.

In the final paper, 'Ethics by Design: Responsible Research & Innovation for AI in the Food Sector', Peter Craigon and colleagues apply three different design methodologies to consider the ethical challenges of data sharing: ideation and speculative scenario development, creation of design fiction objects, and Moral-IT card-based tool. These methods elicited considerable anticipation and engagement with 'real' design fictional artifacts, while supporting an ongoing process of reflection and potential action to mitigate risks and keep informing the responsible development of future TAS.

Message in a bottle

Taken together, the papers in this special issue contribute to the literature on RRI and provide insights for practice. First, the papers make recommendations for tools that facilitate reflection, anticipation and engagement among researchers and stakeholders, in the early-design stages of projects, such as LEGO® SERIOUS® Play, ideation and speculative scenario development, creation of design fiction objects, and Moral-IT cards. These methodologies invite participants to imagine, project and elucidate fictitious but plausible scenarios, effective for identifying and anticipating potential harm and benefit in TAS. Second,

RRI should be an activity that fundamentally changes the ways in which research is designed and conducted. Third, successful RRI ecosystems should contribute to institutional and professional reputation building through honest and transparent stakeholder engagement, respect, commitment to action plans, and responsiveness to issues that may arise with TAS research. Fourth, RRI 'attunement' and considerations in research are not a one-time tick box, they are an ever-ongoing effort, a mindset or attitude towards research and innovation. Finally, whether RRI implementation can and should be monitored, and how this should be measured to understand the impact of RRI on research outcomes, are still open questions that need to be answered.

Elvira Perez Vallejos*, Liz Dowthwaite, Pepita Barnard, Ben Coomber

* Corresponding author.
E-mail address: elvira.perez@nottingham.ac.uk (E.P. Vallejos).