

# Introduction to Responsible Research and Innovation



Dependable and Deployable AI for Robotics 2

Barbara Webb

# Introductions

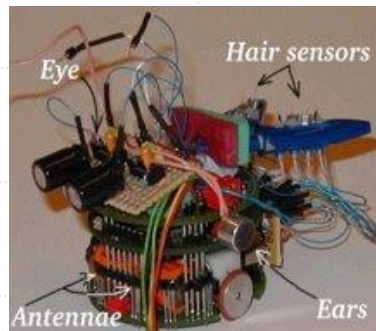
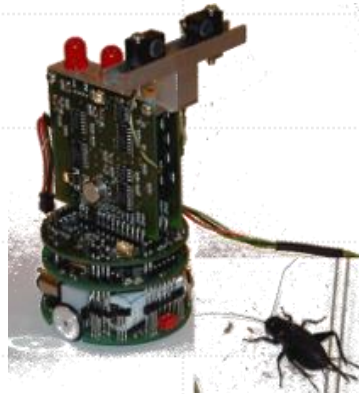
Who am I?

Professor in Biorobotics

I try to understand the mechanisms underlying the efficient and robust behaviour of insects

Focussing on tasks relevant to robotics

And building robots to test mechanistic hypotheses



Who are you?

What previous experience do you have on the topic of responsible research & innovation?

Formal study in your previous degree(s)?

Industry/job experience?

Wider life experience?

# Scope of the course

## Week 1

- What is RRI and how might it apply to your research? Introducing the AREA framework.
- How should RRI issues interact with system design approaches?
- How do regulations, standards and guidelines constrain RI choices?
- You will explore some topical RRI issues in robotics (Thursday) and present results (Friday)

## Week 2

- AREA tools: Engagement
- AREA tools: Anticipation and Reflection
- AREA tools: Action
- You will prepare a draft AREA plan for your research (Thursday) and present it (Friday)
- Assessment: You will refine your AREA plan and submit by Feb 28.





# What is Responsible Research and Innovation?

Many definitions available, but we will use this succinct summary –

**“Aligning research and innovation to the values, needs and expectations of society”**

Note this includes both the process (how you do the research) and the outcome (what the research produces).

It encompasses issues of:

- Ethical acceptability
- Social desirability
- Safety
- Security
- Sustainability

This is not just about driving towards desirable ends but also avoiding undesirable side-effects.

(From D2AIR-1)

# What is Systems Engineering? (INCOSE)

*Systems Engineering (SE) is an interdisciplinary approach and means to enable the realization of successful systems.*

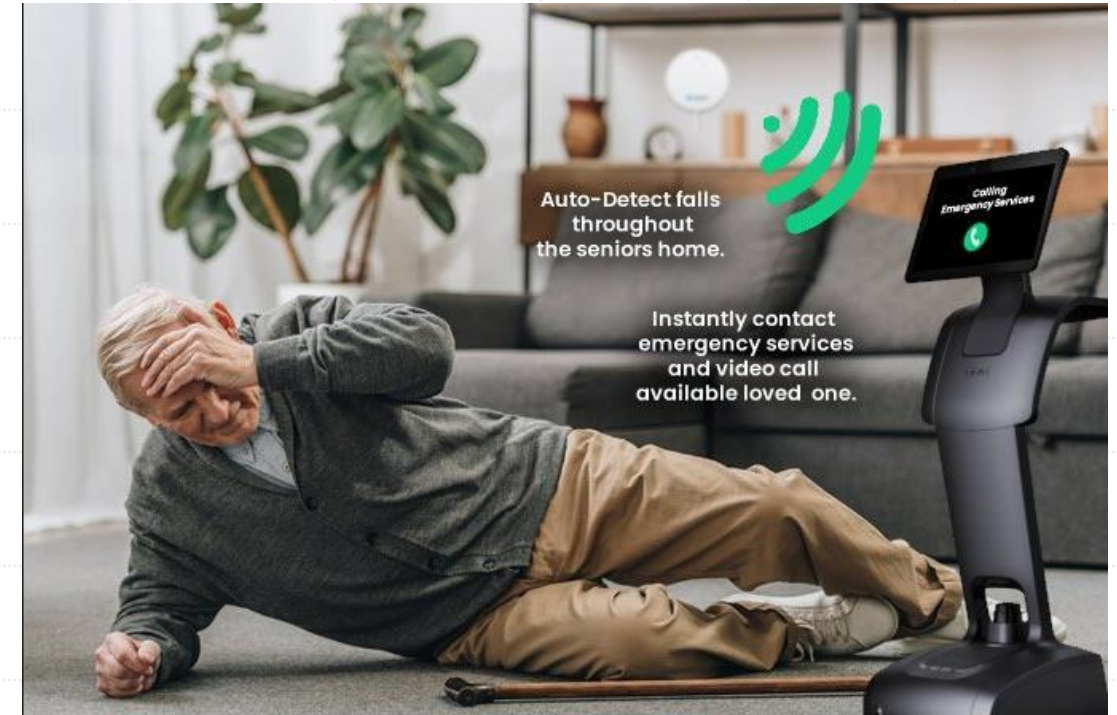
*It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, and then proceeding with design synthesis and system validation while considering the complete problem: operations, cost and schedule, performance, training and support, test, manufacturing, and disposal.*

*SE considers both the business and the technical needs of all customers with the goal of providing a quality product that meets the user needs*

*Responsible SE also considers the values, needs and expectations of wider society, and how these can be met in the design process and outcome.*

# Example

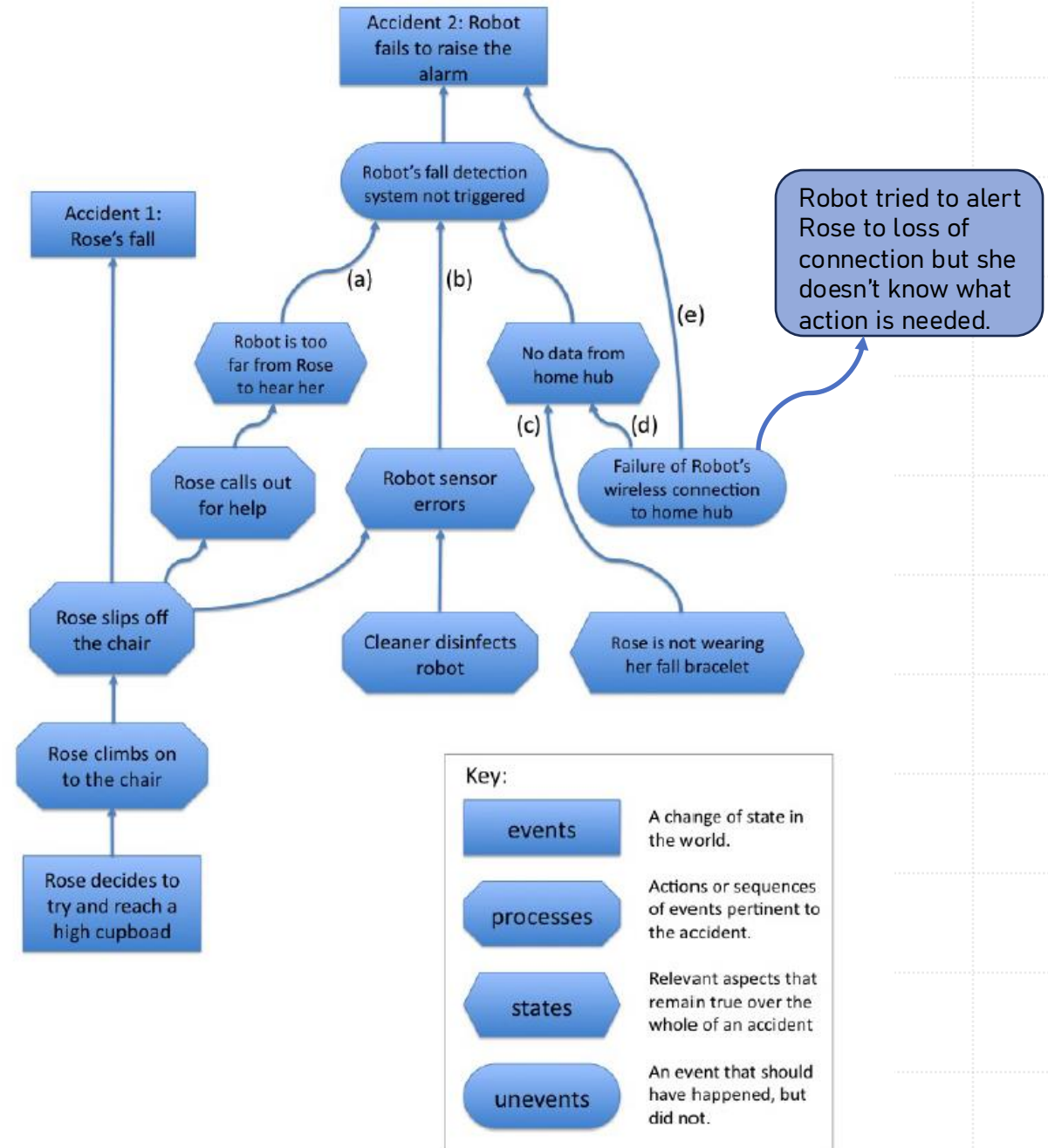
- Robot accident investigation (Winfield et al. 2021)
- Scenario – a carer finds a woman collapsed on the floor in her kitchen while the robot appears to be moving aimlessly around the apartment. It failed to send any alert and could possibly have been responsible for causing the fall. Luckily the woman recovers, but what went wrong?
- Introducing social robotics means human accidents involving robots will increase, due to more interactions, naïve users, messy environments, range of possible harms...
- Many sectors (e.g. aviation, medical) have established accident investigation procedures:
  - Systematic models of safety, e.g. swiss cheese
  - Examine entire sociotechnical system to understand weaknesses in safety defences
- So far there is no well-established procedure for robotics



Winfield, et al. (2021). Robot Accident Investigation: A Case Study in Responsible Robotics. In: Software Engineering for Robotics. Springer, Cham. [https://doi.org/10.1007/978-3-030-66494-7\\_6](https://doi.org/10.1007/978-3-030-66494-7_6)

# Example

- Robot accident investigation (Winfield et al. 2021)
- Proposal:
  - Robot should have 'black box' equivalent, recording recent sensor data, actuation commands, location, high-level AI decisions, robot status.
  - Investigate the whole situation and create a “why-because” graph to determine causal factors
- Bigger picture:
  - Who is on the investigation team?
  - Is even clear how to contact the robot company? Are they obliged to investigate and report?
  - Should there be an agency (e.g. Health and Safety) responsible and do they have the expertise for analysing robot failures?
- Shouldn't these issues have been identified, and ameliorated, in the design process?



# What are some key issues for RRI in the context of AI for Robotics?

- Potentially disruptive technology in variety of sectors
- Complex integrated systems operating in uncertain environments
- Potential for physical interactions with humans including non-expert users
- Existing standards for physical devices and control systems don't address AI and ML issues





# RRI sounds good, but how?

- Codes of practice? <https://www.acm.org/code-of-ethics>
- Manifestos? <https://criticalengineering.org/>
- Refuse to design? <https://direct.mit.edu/books/oa-monograph/4605/chapter-standard/211376/Directions-for-Future-Work-From-TechWontBuildIt-to>

# The AREA framework

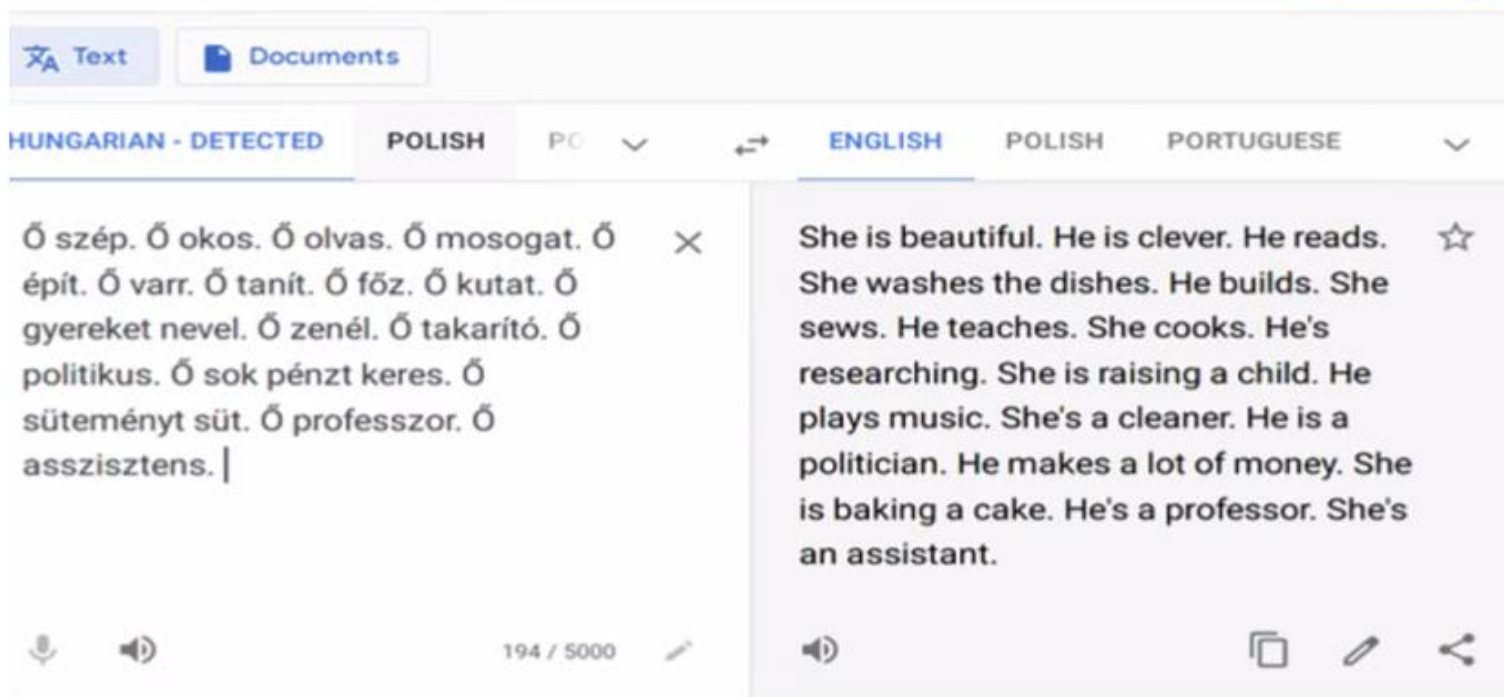
- **Anticipate** outcomes
  - **R**eflect on motivation, processes and products
  - **E**ngage with stakeholders
  - **A**ct responsively
- 
- See <https://rai.ac.uk/toolkits/rri-toolkit/>

Note an equivalent four-part framework often cited is from Owen et al 2013,  
(<https://doi.org/10.1002/9781118551424.ch2>)

- Anticipation
- Inclusion
- Reflexivity
- Responsiveness

# The AREA framework

- Anticipate outcomes
- Both intended and unintended
- Both positive and negative
- An example of an unanticipated outcome of using LLM translation: Dora Vargha (2021) Twitter, <https://twitter.com/DoraVargha/status/1373211762108076034>



Who is responsible?



# The AREA framework

- **Anticipate outcomes**
  - Both intended and unintended
  - Both positive and negative
- **Reflect on motivation, processes and products**
  - Specifically this involves looking at (and potentially challenging) your own motivations, your perspective on the problem, why you are taking a particular approach.
- **Engage with stakeholders**
  - Need to think beyond the immediate customer to wider society
- **Act responsively: the previous processes should result in actions.**
  - This could involve gathering more information and committing to return to previous steps





# Exercise:

Considering your own research plans, brainstorm answers to the following questions:

[Purpose] Why do this? Is it potentially controversial or disruptive?

[People] Who will be affected and how? Directly and indirectly? Benefitted or harmed? Have they had, or will they have, any input?

[Product] What are the potential uses, including unintended? What are the environmental impacts of production, use and disposal?

[Process] How are any of these issues going to affect your project plan?

# Exercise

- Role-play
- Imagine you are a PI leading a robot research team. You are being interviewed by a group of journalists.
- 1) The media have exaggerated the positive application potential of your research.
- 2) Your research is of high theoretical interest but unlikely to have much application. The media want to know why the tax-payer should be funding it.
- 3) The media have latched onto a potential harmful application of your research.
- Debrief: RRI includes honest communication with society