"What Could Possibly Go Wrong?" Logging HRI Data for Robot Accident Investigations

Katie Winkle
katie.winkle@brl.ac.uk
Bristol Robotics Laboratory
Bristol, U.K.

Marina Jirotka
marina.jirotka@cs.ox.ac.uk
University of Oxford
Oxford, U.K.

Ulrik Lyngs
ulrik.lyngs@cs.ox.ac.uk
University of Oxford
Oxford, U.K.

Carl Macrae
carl.macrae@nottingham.ac.uk
University of Nottingham
Nottingham, U.K.

Helena Webb
helena.webb@cs.ox.ac.uk
University of Oxford
Oxford, U.K.

Alan Winfield
alan.winfield@brl.ac.uk
Bristol Robotics Laboratory
Bristol, U.K.

ABSTRACT

This abstract presents proposed experimental work to consider what might be required for an 'ethical black box', essentially a robot data recorder, to inform robot accident investigation processes and the implications for HRI.

CCS CONCEPTS

• **Information systems → Data management systems**; • **Applied computing → Law, social and behavioral sciences**.

KEYWORDS

ethical black box; data; robot accident investigation; assistive human robot interaction

1 INTRODUCTION

RoboTIPS is a 5-year project considering the development of responsible robots for the digital economy. Key aims of the project are to (i) develop the specification for an Ethical Black Box (EBB) for social robots that autonomously interact with people and (ii) demonstrate how an EBB might be utilised in the context of robot accident investigation. This abstract presents initial design work considering what might be required for an 'ethical black box' in the context of an HRI scenario (assisted living) and a draft experimental procedure for a mock HRI accident and investigation process in which that can be tested.

1.1 An Ethical Black Box for Robots

Winfield and Jirotka recently proposed the concept of an 'ethical black box' (EBB), the equivalent of a Flight Data Recorder for robots and autonomous systems [2]. The proposed EBB would continuously record sensor and internal state data, with the objectives of (i) informing robot accident investigations, specifically in establishing accountability and responsibility, and (ii) affording a level of transparency which positively impacts public trust of robots and autonomous systems.

2 CONSTRUCTING A MOCK HRI ACCIDENT & INVESTIGATION SCENARIO

To our knowledge this is the first work to consider processes for HRI accident investigation, such that how to design and experimentally test an EBB in the context of such processes is an open research question. To address this, the project team, consisting of researchers from HRI, human-centred computing and experimental psychology undertook a co-design session in-situ at the Assisted Living Studio of the Bristol Robotics Laboratory, set-up as per Figure 1. The following scenario, proposed by Winfield as one exemplar demonstrating the need for robot/AI accident investigation (see [1] for full details) was used to introduce the scene and form the basis of the proposed accident scenario.

"Your elderly mother, or grandmother, has an assisted living robot to help her live independently at home...one afternoon you get a call from a neighbour who has called round and sees your grandmother collapsed on the floor. When the paramedics arrive they find the robot wandering around apparently aimlessly. One of its functions is to call for help if your grandmother stops moving, but it seems that the robot failed to do this...the doctors find bruising on her legs, consistent with the robot running into them. Not surprisingly you want to know what happened: did the robot cause the accident? Or maybe it didn't but made matters worse, and why did it fail to raise the alarm?"

Topics of discussion, designed to inform initial EBB design and mock accident/investigation experimental procedure included:

- Where is the *boundary* of an assisted living robot that is likely to rely on/utilise data from a connected smart home infrastructure?
- Who are the *witnesses* in a robot accident scenario?
- What data could be supplied to/would be required by an accident investigation team?

Figure 1: Assisted living studio set-up for accident and investigation process co-design session, resulting in the proposed experimental procedure given in Section 2.1.

Mock Accident: T_0 EBB Data Generation
The accident will be physically enacted in the assisted living studio in order to generate realistic EBB data records.
Mock Accident: Witness Testimony Generation
The *neighbour* and then the *paramedic* will be shown to the scene and allowed unprompted and unrestricted interaction with the *user* who will then be *removed to the hospital*. Written testimony will be collected at this point, and the scene will be photographed.
Investigation: Briefing Accident Investigation Team
The accident investigation team will be assembled, briefed on the situation and presented with (i) the EBB data available for before, during and immediately after T_0 according to the EBB specification and (ii) the written witness testimonies.
Investigation: Process & Presentation of Findings
The accident investigation team will be hosted on-site to conduct a 1-day investigation process and to report back preliminary findings concerning what they believe to have happened and why. This investigation process will be observed by the project team, with video and audio data recorded for thematic analysis as part of the experimental data collection process.

Table 1: Experimental protocol for mock HRI accident and related investigation process.

- What domains/expertise should be represented on a robot accident investigation team?
- (How) can existing investigation processes/accident models, drawn from e.g. the aeronautical and healthcare industries, be used to inform an initial robot accident investigation procedure?

These discussions resulted in the following experimental procedure. A number of research questions specifically pertinent to HRI were also identified, as presented in Section 2.2.

2.1 Proposed Experimental Procedure

We have designed an experimental procedure covering both the mock accident and its resultant investigation process, key elements of which are presented in Table 1. The mock accident will be conducted in the assisted living studio as pictured in Figure 1. Participants and actors referred to in the experiment protocol are as follows:

User - An actor will be employed and briefed to the play the role of the robot user: an 85 year old lady who lives alone in a 1 bedroom flat located within a retirement village.

Witness 1: Neighbour - A naive, non-roboticist be will recruited to play the role of the user's neighbour. They will be briefed that they are taking part in a roleplay regarding the use of robots in assisted living, and their role is to call in on their neighbour who uses such a robot. They will not be briefed in advance regarding the accident.

Witness 2: Paramedic - A naive, non-robotocist healthcare professional will be recruited to play the role of a paramedic, again briefed regarding the experiment being a roleplay exploring assisted living robots. They will be asked to attend the *user* on the scene as they would a normal emergency callout.

Accident Investigation Team - The accident investigation team will be selected to represent expertise in/interests of the domains of accident investigation, assisted living robotics and supported housing/healthcare domains. The investigatory team will not include any of the project researchers.

An exact specification for the EBB is still under development but data collected and recorded is likely to include: any smart home/auxiliary sensor data regarding user position and status; robot video footage; logs of human-robot interaction with voice recordings and internal robot state information (e.g. operation mode, perceived user status etc.).

2.2 Pertinent HRI Research Questions

In considering the above accident scenario, it became clear that any user interactions engaged in by the robot would be of serious interest to the accident investigation team, and likely crucial to establishing what happened/why. This raises the following research questions concerning human robot interaction and its documentation 'in-the-wild':

(1) To what extent do human robot interactions need to be *logged*, and how, in order to satisfactorily inform accident investigation processes? What user interaction and robot *sense-control-act* data could be recorded/stored?
(2) What do 'first on scene' witness testimonies tell us about human perception and expectations of social robots where an accident has occurred?

3 CONCLUSION

We present proposed experimental work to investigate use of an EBB to inform accident investigation in an applied HRI context. Future work will iterate both the EBB design and investigative process across different scenarios in order to contribute towards the development of responsible robots.

REFERENCES

[1] 2019. Alan Winfield's Web Log: What's the worst that could happen? Why we need robot/AI accident investigation.

[2] Alan FT Winfield and Marina Jirotka. 2017. The case for an ethical black box. In *Annual Conference Towards Autonomous Robotic Systems*. Springer, 262–273.