In-the-Wild Failures in a Long-Term HRI Deployment

Francesco Del Duchetto
L-CAS, School of Computer Science
University of Lincoln
Lincoln, UK
fdelduchetto@lincoln.ac.uk

Ayse Kucukyilmaz
School of Computer Science
University of Nottingham
Nottingham, UK
ayse.kucukyilmaz@nottingham.ac.uk

Marc Hanheide
L-CAS, School of Computer Science
University of Lincoln
Lincoln, UK
mhanheide@lincoln.ac.uk

***Abstract*—Failures are typical in robotics deployments "in-the-wild", especially when robots perform their functions within social human spaces. This paper reports on the failures of an autonomous social robot called Lindsey, which has been used in a public museum for several years, covering over 1300 kilometres through its deployment. We present an analysis of distinctive failures observed during the deployment and focusing on those cases where the robot can leverage human help to resolve the problem situation. A final discussion outlines future research directions needed to ensure robots are equipped with adequate resources to detect and appropriately deal with failures requiring a human-in-the-loop approach.**

I. Introduction

In recent years, various successes in robot navigation, perception and planning enabled long-term deployments of robot systems in a somewhat autonomous fashion. Robots are more and more able to operate in dynamic environments while perceiving their surroundings and the other entities in them. Although robots can be embedded with state-of-the-art manipulation, learning, and visual perception abilities, they have not yet reached a level of autonomy which guarantees failure-safe operation [7]. As such, even the longest running deployments with state-of-the-art robotic systems show that unexpected situations and failures are virtually impossible to avoid [8].

As autonomous failure detection and recovery in-the-wild is a hard research question, robotic systems of our day rely on frequent expert interventions when failures occur. These interventions may involve fixing a bug that caused the issue in the first place, re-designing the system, or modifying and replacing hardware to increase the robot perception and/or computing power. Such resolutions are offline and require expert programmer/engineer interventions. On the other hand, in many cases, the unexpected situations can be resolved by non-expert humans if the robot is able to detect the issue and instruct humans on how to solve it. For example, if a robot is unable to navigate because its wheels are stuck on a carpet, it can ask for help from a nearby human for being freed away from the obstacle, as is common practice with many robotic vacuum cleaners used in domestic settings.

This work is partially funded by the Lincolnshire County Council, and EPSRC under grant IDs EP/V00784X/1 (TAS-Hub: TARICS) and EP/T022493/1 (Horizon: FAILSAFE)

Although performing tasks in human environments introduces challenges for robots, there is a benefit to being among humans: robots can always rely on help from nearby humans to resolve difficult situations [12].

In this paper, we report on the failures that occurred over the three-year-long deployment of Lindsey, a tour guide robot in an archaeological museum [1], where the robot was programmed to actively seek human interventions as a recovery or mitigating strategy for most failure cases. In analysing these failures and recoveries, we attempt to inform future deployments of autonomous social robots in-the-wild and suggest strategies that can be employed to integrate the users into the recovery procedure effectively.

II. Related Work

Seminal studies in long-term deployments have initially focused on robustness to allow autonomous operations, identifying the interaction with humans as a necessity for recovering from failures and performing tasks that the robot was not able to [1], [11]. Indeed, several ad-hoc recovery strategies to recover from failures have been proposed in the literature, including within the STRANDS project [8], which featured hard-coded autonomous recovery behaviours. These range from a simple (i) *wait and retry* behaviour, which clears the local cost-map and then re-issues the navigation command, over a (ii) *backtrack* behaviour, which reversed the last N motion commands, covering several seconds of operation, sent to the robot to return to a previous position, up to an (iii) *interactive help seeking* behaviour, in which the robot would ask any human in its surrounding –verbally and by screen display– to push it to a free area. Earlier analysis in the STRANDS experiment [8] showed that around one third of all robotic failures required eventual human help as a last resort, after all autonomous behaviours had failed. These examples underline the usefulness of deploying robots to perform tasks in human environments, even when their tasks do not require interacting with humans, where the robots can always rely on the help from nearby humans to resolve difficult situations [12]. Such human interactions are precious and rare, and previous work builds on interactively learning

[1] https://www.lincolnmuseum.co.uk/robot-at-lincoln-museum

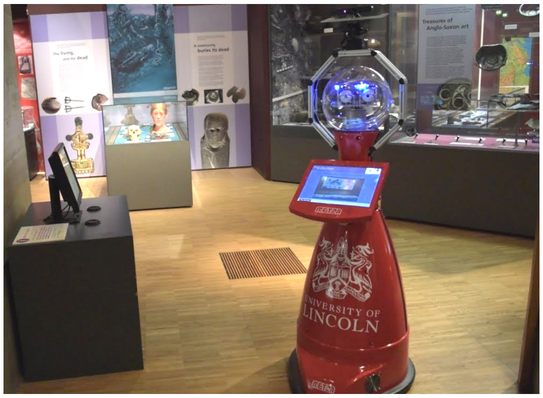

Fig. 1. Lindsey, the mobile robot deployed as a tour guide in the Lincoln Museum.

from human demonstrations as an invaluable tool to reduce the robot reliance on humans over time [7]. Additionally, in a scenario similar to the one presented here, Wang et al. [14] reported extensive hardware, software and interaction failures over their one-month deployment with a robot tour guide system.

Previous work has studied failures in human-robot interaction deployments by proposing different taxonomies for their categorisation, in particular focusing on the user perception of the failures and recoveries. For example, it has been studied how robotics failures affect trust and the mitigation strategies that can be employed thereafter [10], [13].

In this paper, we report on the failures that have occurred in our long-term museum deployment and we focus on those were the users could provide assistance to solve the issue or those that directly involves the users in the interactions. Our aims are to understand what strategies, whether direct or by using learning approaches, can be used to increase the robot autonomy by leveraging the users help in long-term scenarios.

III. In-the-Wild Failures and Recoveries

This short article reports on the failures that have happened during the long-term deployment of a tour guide robot, called Lindsey and shown in Figure 1. The deployment started in 2018 and is ongoing to date, having totalled more than 446 days of autonomous operation (excluding the periods of interruptions caused by the COVID-19 lockdowns and malfunctions) and navigated more than 1300 kilometres. The robot is available daily to the visitors of the museum, who can start different interactive tasks with the robot, such as a thematic guided tour where the robot guides them to several items and explains what they are. More details about the robot system and the interactive tasks are reported in our previous work [5]. The failures reported in this article were collected in the period between January 2019 and October 2022.

In the deployment presented in this work, most errors and failures happen during navigation because it is the only action performed requiring the robot's physical movement. By being potentially damaging to people and the environment, there are multiple ways the action can fail as a result of safety measures put in place. In addition to navigation failures, we also consider social failures, as in the failures of the robot to behave in a socially coherent way with the users.

The following failures, each with its own recovery strategy, have been considered and reported in this article:

1) *emergency button pressed*: pressing an emergency button on the robot cuts the power to the motors, making the robot unable to move. In order to re-activate the motors, the emergency button must be manually released. If they are pressed when the robot starts a navigation action, it actively asks the users to release the buttons.
2) *bumper pressed*: when the bumper on the base of the robot detects a collision, a software node blocks the robot motors. The motor block can be re-activated on-demand by software. After a collision is detected and the motors are blocked, the robot asks the users to be slightly pushed to signal that it can continue to navigate. Therefore, if a push is detected (as a change in the robot's odometry), the motor block is released.
3) *navigation failure*: the navigation planner is not able to generate a viable path in the topological map because the robot is stuck close to some obstacles. In this case, the robot immediately stops its navigation action and interactively asks the user to be dragged away from the obstacle. After being moved, it restarts the navigation action.
4) *localisation failure*: the localisation algorithm is not able to correctly identify the location of the robot in the environment. This failure can be recovered by making the robot move around in an uncluttered environment to recover the correct localisation, or by returning the robot to its charging station (at which point the localisation is automatically reset to the correct position). The robot asks the users to contact a member of the museum's staff to be moved back onto its charging station.
5) *navigating into a prohibited area*: the museum staff can select specific areas of the gallery to block the robot from navigating in them. This feature is helpful in several situations, such as when there is a school visiting, when there are teachers who do not want to be distracted by the robot, or for maintenance reasons. The area blocking action can happen at any moment while the robot is operating and has an immediate effect; this may cause the robot to inadvertently navigate to such areas or to find itself already in it at times. When this happens, to ensure compliance with the blocking request, the navigation is halted immediately with the robot asking the users to call a member of staff to be pushed outside the blocked area.
6) *users walking away*: this failure happens when the users involved in an on-going interaction with the robot walk away from the interaction before it is finished. The robot can detect this failure by asking the users to confirm, at specific points during the tasks, whether they want to execute a certain action (for example, go to the next

TABLE I
FAILURES AND RECOVERIES SUCCESS.

Failure	Helped Successful	Helped Unsuccessful	Non helped	Total
1	166 (48.82%)	0 (0.00%)	174 (51.18%)	340
2	3220 (91.35%)	0 (0.00%)	305 (8.65%)	3525
3	706 (43.29%)	379 (23.24%)	546 (33.48%)	1631
4	7 (5.79%)	32 (26.45%)	82 (67.77%)	121
5	0 (0.00%)	15 (34.09%)	29 (65.91%)	44
6	N/A	N/A	N/A	470
7	N/A	N/A	N/A	5194
All failures				**11325**

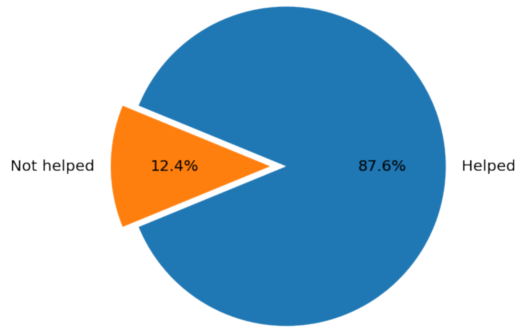

Fig. 2. The rate of *emergency button pressed* and *bumper pressed* failures that were helped by the users when the robot asked them to.

exhibit in the tour). If the robot does not receive a response after one minute, it assumes that the users have left the tour, and hence terminates the interactive task. No recovery strategy is implemented to try and bring the users back in the interaction.

7) *users stop the interaction*: this failure happens when the users willingly stop the interaction with the robot before it is finished. No recovery was implemented for this class of social failure; the robot acknowledges the users' request to stop the interaction and farewells them.

IV. Analysis of Failure Recoveries

Here, we report on the amount of failures that have happened during the long-term deployment, divided into the different failure types described above. In analysing these failures, we report on the following amount: a) the rate of failures in which the users have actively sought to help the robot; b) the rate of failures that have been successfully resolved by a recovery performed by the user.

Because of the nature of the failure types, a recovery procedure was not devised and implemented for the *users walking away* and *users stop the interaction* failures. For the other failure categories, a helped failure was further classified as successful or unsuccessful by using the following criteria:

- *emergency button pressed* and *bumper pressed*: always successful, because the recovery requires a simple action that, when executed, immediately allows the robot to resume its operations.

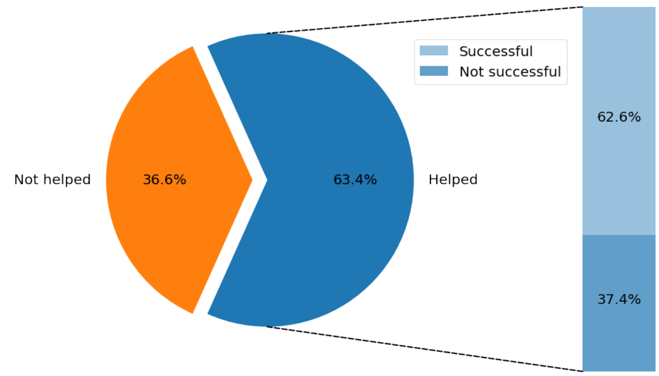

Fig. 3. The rate of *navigation failure*, *localisation failure* and *navigating into a prohibited area* failures that were helped by the users when the robot asked them to. Out of all the helped ones, it shows the portion that turned out to lead the robot to recover successfully.

- *navigation failure*: the recovery is considered successful when the robot is able to successfully reach the navigation goal afterwards.
- *localisation failure*: the recovery is successful when the robot localisation improves to a level that enables autonomous navigation before the end of the task.
- *navigating into a prohibited area*: the recovery is considered successful whenever the robot can successfully navigate outside a prohibited area before the end of the task, implying that the robot has been pushed outside the area.

Table I shows the total absolute numbers of the failures described in Section III with the portions that have been helped/non-helped and, among the helped ones, those that were successful/unsuccessful. In Figure 2, we show the rate of failures 1 and 2 that have been helped versus those that haven't. In Figure 3, we show instead the rate of failures 3, 4 and 5 that were helped/non-helped, and those that were successfully recovered afterwards.

From the data reported here, we can observe that some failures are inherently more easy to recover from by exploiting the human presence in the environment. For example, for the failures reported in Figure 2 users were willing to help the robot in more than 85% of the cases. By contrast, the remaining failures (excluding those that cannot be helped) were helped just over 63% of the time. One possible explanation for this difference is that, in the first case, the recovery procedure requested is more easily understandable as it just requires releasing a button or slightly pushing the robot, while, in the second case, a more effortful recovery needs to be executed by the users. Requesting user recoveries for the second group of failures is still a valid policy since the chances of the users helping and of the recovery being successful outweigh that of the users not helping at all; however, the recoveries received in this situation are more scarce and precious.

V. Challenges and Opportunities

Here, we review a handful of challenges and opportunities for future research that are particular to failure management with service robots.

A. A wider taxonomy of failures

Although validation and verification are well-studied for software and industrial robotics applications, there are only a few studies which investigate failures in human-robot interactions [3]. In addition to technical failures, robots can also fail socially even though they are functionally error-free [3], and any perceived failure makes the robot seem less capable, lowers users' trust, and can make people reluctant to use the service again [2], [4]. This issue is evident in our present deployment considering the large numbers of *users walking away* and *users stop the interaction* failures, for which no recovery procedure was found so far. There is a need for multidisciplinary studies, bringing together different stakeholders, to develop novel socio-technological theories for human perception of robotic failures. Such an effort would promote the development of human-robot interaction tools for user-centred failure handling strategies.

B. Intuitive mechanisms for expressing failure situations

For non-expert users, it could be difficult even to reliably identify if a robot is functioning properly or failing, and to choose how to act if they perceive a failure. On top of this, cognitive, psychological, and social determinants that should impact the design of communication and failure mitigation strategies are not well-studied [9]. These shortcomings limit the design of effective management strategies for faulty and unexpected behaviour observed by untrained users. As explainability of AI is gaining interest across research fields, there is great potential to design and develop robotic behaviours that would enable easier failure identification and resolution.

C. Interactive and iterative learning systems

In a previous work [7], we proposed a two-layer cascaded learning approach, where an "ask-for-help" paradigm is implemented on Lindsey. In this paradigm, the robot gathers human demonstrations when necessary, to incrementally learn local navigation strategies to handle the failure when it happens. To demonstrate the use of this model, we identified two failure scenarios, where the global navigation typically fails [2]. This approach was useful to progressively increase the robot's competencies in detecting failures and recovering from them. Differently, in [6], we implemented an in-situ reinforcement learning algorithm to adapt the robot's behaviour based on minimising the chances of the users stopping or abandoning the interaction with the robot. This approach can be seen as a machine learning solution aimed at limiting the social failures in human-robot interactions. Starting from the assumption that failures cannot be eliminated or recovered altogether in in-the-wild deployments, there is a great opportunity for integrating machine learning methods in failure recovery and avoidance frameworks to enable robots to continually learn to improve their autonomy.

[2]The global navigation system on Lindsey is based on the *move_base* package from the ROS navigation stack.

VI. Conclusions

In this short paper, we reported on the failures that have been observed over a 3-years long deployment of a social robot in a public museum. By analysing the "human-in-the-loop" recoveries for such failures, it can be observed that the willingness of providing help and the effectiveness of the help given varies among different failures. Based on these results and on previous findings in the literature, we provide suggestions for future research needed to ensure robot autonomy in human-populated environments in the wild.

References

[1] Biswas, J., Veloso, M.: The 1,000-km challenge: Insights and quantitative and qualitative results. IEEE Intelligent Systems **31**(3), 86–96 (2016)

[2] Brooks, D.J.: A human-centric approach to autonomous robot failures. Ph.D. thesis, University of Massachusetts Lowell (2017)

[3] Brooks, D.J., Begum, M., Yanco, H.A.: Analysis of reactions towards failures and recovery strategies for autonomous robots. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 487–492. IEEE (2016)

[4] Cha, E., Dragan, A.D., Srinivasa, S.S.: Perceived robot capability. In: 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 541–548. IEEE (2015)

[5] Del Duchetto, F., Baxter, P., Hanheide, M.: Lindsey the Tour Guide Robot - Usage Patterns in a Museum Long-Term Deployment. In: International Conference on Robot & Human Interactive Communication (RO-MAN). IEEE, New Delhi (2019)

[6] Del Duchetto, F., Hanheide, M.: Learning on the Job: Long-Term Behavioural Adaptation in Human-Robot Interactions. IEEE Robotics and Automation Letters **7**(3), 6934–6941 (may 2022). https://doi.org/10.1109/lra.2022.3178807

[7] Del Duchetto, F., Kucukyilmaz, A., Iocchi, L., Hanheide, M.: Do not make the same mistakes again and again: Learning local recovery policies for navigation from human demonstrations. IEEE Robotics and Automation Letters **3**(4), 4084–4091 (2018)

[8] Hawes, N., Burbridge, C., Jovan, F., Kunze, L., Lacerda, B., Mudrova, L., Young, J., Wyatt, J., Hebesberger, D., Kortner, T., et al.: The strands project: Long-term autonomy in everyday environments. IEEE Robotics & Automation Magazine **24**(3), 146–156 (2017)

[9] Honig, S., Oron-Gilad, T.: Understanding and resolving failures in human-robot interaction: Literature review and model development. Frontiers in psychology **9**, 861 (2018)

[10] Marinaccio, K., Kohn, S., Parasuraman, R., Visser, E.J.D.: A framework for rebuilding trust in social automation across health-care domains. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care **4**(1), 201–205 (2015). https://doi.org/10.1177/2327857915041036, https://doi.org/10.1177/2327857915041036

[11] Meeussen, W., Marder-Eppstein, E., Watts, K., Gerkey, B.P.: Long term autonomy in office environments. In: ALONE Workshop, In Proceedings of Robotics: Science and Systems (RSS'11), Los Angeles, USA (2011)

[12] Rosenthal, S., Biswas, J., Veloso, M.: An effective personal mobile robot agent through symbiotic human-robot interaction. In: Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS. vol. 2, pp. 915–922 (2010). https://doi.org/10.5555/1838206, www.ifaamas.org

[13] Tolmeijer, S., Weiss, A., Hanheide, M., Lindner, F., Powers, T.M., Dixon, C., Tielman, M.L.: Taxonomy of trust-relevant failures and mitigation strategies. ACM/IEEE International Conference on Human-Robot Interaction pp. 3–12 (3 2020). https://doi.org/10.1145/3319502.3374793

[14] Wang, S., Christensen, H.I.: TritonBot: First Lessons Learned from Deployment of a Long-Term Autonomy Tour Guide Robot. In: RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication. pp. 158–165 (2018). https://doi.org/10.1109/ROMAN.2018.8525845, https://github.com/CogRob/