Lindsey the Tour Guide Robot: Adaptive Long-Term Autonomy in Social Environments

Francesco Del Duchetto

Director of Study Prof. Marc Hanheide
Second Supervisor Dr. Paul Baxter

UNIVERSITY OF LINCOLN
School of Computer Science
College of Science

A thesis submitted in partial fulfilment of the requirements of the University of Lincoln for the Degree of **Doctor of Philosophy**

This research programme was partially sponsored by the Lincolnshire County Council

June 2023

Acknowledgements

I would like to thank my director of study, Prof Marc Hanheide, for his continuous support during my PhD studies. He has always been able to direct me to the right path when I was uncertain while giving me the space to form and explore my own ideas. I am grateful for all the discussions, the suggestions and the opportunities he has given me over the years, and still to date. I would like to thank also my second supervisor, Dr Paul Baxter, for the fruitful insightful discussions and for showing me different perspectives on my work. I would also like to acknowledge my examiners, Prof Simon Parsons and Dr Mary Ellen Foster, for reviewing my work and providing their honest comments that have improved the quality of this thesis.

I am most grateful to my parents for their support, for always being there for me and for providing an environment that allowed me to find my path in life. To my brother and sister, for the fun times together and their support; to my grandparents, for teaching me many things in life, and to the rest of my family that have always stood by my side. I want to thank Michele for always being my best support, during the good and the bad times, for the joyful moments and for showing me a different perspective on things. Last but not least, I want to thank all my colleagues and friends from L-CAS, for creating a stimulating environment that made me want to choose Lincoln over and over, and for my other friends, in Lincoln and around the world, on which I can always rely for a friendly chat.

Abstract

THIS project proposes a framework for online adaptation of robot behaviours deployed autonomously in social settings with the goal of increasing the overall users' engagement during the interactions.

One of the most critical aspects to address for robots deployed in "the real world" is the necessity of interacting with people, whether intentionally or not. Interacting with people requires a wide range of capabilities, from perceiving the different people's intentions and emotional states to generating appropriate behaviours for the specific context of the interaction. Moreover, it requires that robots learn and adapt from experience while interacting with their users. In this project, a mobile robot is embedded in a long-term study in a public museum. The robot has been deployed for more than a year, to date, as an autonomous tour guide to the museum's visitors, with its tasks being guiding people to the position of various exhibits and giving a description of each item. The long-term scenario allows studying how people interact with a robot in an unconstrained setting and give the opportunity of improving the current state-of-the-art robotics autonomy in a social setting.

The initial data collection shows that users' engagement during the robotised tours steeply declines after the initial moments of the interaction. The first main contribution of this project is to investigate whether it is possible to automatically assess the users' engagement from the robot point-of-view during the interactions. A dataset of robot ego-centric videos was collected and manually annotated by independent coders with continuous engagement values. From it, an end-to-end regression model was trained to predict engagement from the robot point of view from a single camera. Experimental evaluation shows that the model accurately estimates the engagement level of people during an interaction, even in diverse environments and with different robots.

Once the robot can detect the engagement state of users during the interactions, it can potentially plan tangential behaviours to influence the users' attentional state itself. The second contribution of this work is devising an online reinforcement learning algorithm that allows the robot to adapt its behaviour online from the feedback obtained during the interactions. The feedback is obtained from users' engagement values estimated from the robot head camera. In the experimental evaluation, the robot delivers the usual tours to the users with the difference that

the choice of some actions is left to the adaptive learning algorithm. Results show that after a few months of exploration, the robot successfully learns a policy that leads people to stay in the interaction for longer.

Contents

List of Figures

List of Tables

Acronyms

AI Artificial Intelligence. 15

ANN Artificial Neural Network. 52–54

CNN Convolutional Neural Network. 53, 54, 62, 64

DL Deep Learning. 51, 52, 60, 70, 71, 103

FC Fully Connected. 62

GPU Graphics Processing Unit. 40

GUI Graphical User Interface. 23

HRI Human Robot Interaction. 4, 46, 48, 50, 63, 78, 84, 99

LSTM Long-Short Term Memory. 54, 62, 64, 70

LTA Long-Term Autonomy. 2, 8, 10, 15, 33

MDP Markov Decision Processes. 78, 79

ML Machine Learning. 73

PNP Petri Net Plans. 19, 27, 31, 82, 83

RL Reinforcement Learning. 3, 5, 11, 15, 45, 46, 70, 73, 74, 77, 78, 81–84, 86, 96, 102

ROS Robot Operating System. 5, 18, 19, 23, 63

UCBVI Upper Confidence Bound Value Iteration. 81, 87, 96, 105

VI Value Iteration. 80, 81

CHAPTER 1

Introduction

Having robots operating among humans and helping them in their daily lives is a goal that is still far from becoming a reality. Although many successful robots are currently deployed in the real world, most of their capabilities usually involve vacuum cleaning your house, inspecting industrial sites, spraying agricultural fields or moving items in warehouses. The commonality of such applications is that humans are not involved in the robot's operations; they are only occasionally around the robot while going on with their daily activities in a way that their presence can be ignored from the robot's perspective and modelled as environmental noise in the observations (e.g. an obstacle to avoid on the path to reach a location). Differently, those applications where the robot operations intrinsically require some level of interaction with humans still have to produce a product that is, as of today, successfully operating autonomously outside the research labs. A recent paper from Tulli et al. [94] analyses why virtually all social robot business initiatives so far have failed to generate a successful product, despite the high expectations from research successes and consumer hype. It suggests that there are many technical, economic and ethical challenges for which we still do not have a clear solution. In fact, in a recent interview[1] various industry leaders from the companies developing such robot technologies explain that managing the user's expectations and modelling their social responses to robot behaviours are amongst the most challenging things to get right when designing a social robot to be deployed in the real world.

With the many challenges come many opportunities to improve the state-of-the-art of social robot autonomy; the studies presented in this thesis are distinct efforts in that direction. Taken together, they are parts of a general framework that allows a robot to interact socially with users in a museum scenario where it is tasked to provide guided tours to visitors.

[1] Why Building Social Robots Is Much Harder Than You Think. (2020, September 16). Retrieved from https://www.topbots.com/building-social-robots-jibo-anki-cozmo-much-harder-think

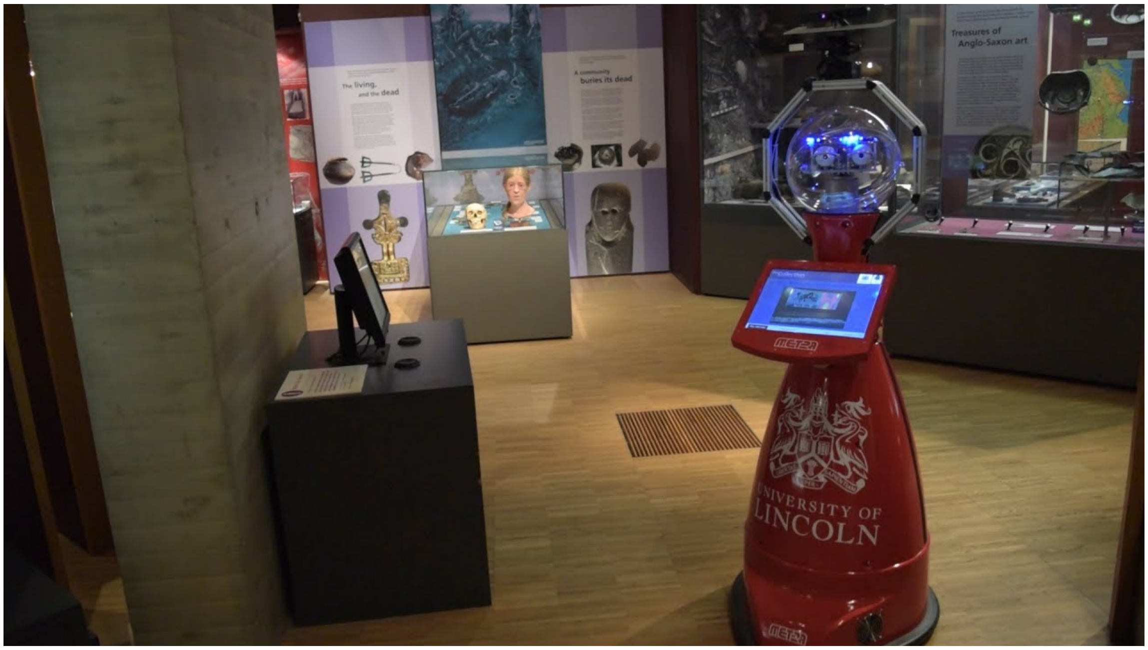


Figure 1.1: Lindsey the museum guide robot deployed in the archaeological gallery.

1.1 Lindsey: the Tour Guide Robot

The research project that informs this thesis is a collaboration between the University of Lincoln and The Collection Museum[2] with the goal of deploying an autonomous robot – Lindsey – to engage with the museum visitors and inform them about the local archaeology. Figure 1.1 shows Lindsey navigating in the gallery.

The museum scenario is a public space openly accessible to anyone in which both the users and the robot can move freely around. Lindsey is meant to be a permanent "museum guide" in the gallery – guiding visitors to various items displayed and describing what they are – being available at all times (during the museum opening hours) and for a deployment lasting years to date. The requirement of long-term deployment in such a public space leads to the first objective of this project:

> **Objective 1**: Deployment of an autonomous robot to provide guided tours to visitors in a public museum. The robot framework must be robust to the point that it enables Long-Term Autonomy (LTA) requiring minimum in-situ assistance from experts.

Guided tours in museums involve several actions to be performed by the robot over an interaction that can last various minutes – the robot needs to describe the general theme of the tour, guide visitors to the items featured in it and give descriptions for each. Moreover, the users do not know what to expect from the first encounter with the robot, not being informed about its functions or how to behave

[2] https://www.thecollectionmuseum.com/

with it. Therefore, from the robot's perspective, starting and sustaining the interaction is a challenge. Based on previous work, I anticipate that:

> **Hypothesis 1**: Robot dynamic behaviours are more effective than static behaviours at maintaining users' engagement throughout the interactions.

There is, therefore, the need to study how people interact with Lindsey:

> **Objective 2**: Analysis of the initial deployment –with a static tour– specifically focused on observing the visitors' engagement with the technology during the guided tours.

Informed by the deployment analysis, there is the possibility of giving Lindsey the ability to automatically detect engagement during the interactions. Based on the observation that engagement is a high-level concept which can be rather intuitive to assess for humans but inherently difficult to formalise into a simple computational model, I hypothesise that:

> **Hypothesis 2**: Humans can intuitively and holistically assess the engagement of the people interacting with them, being able to give a continuous engagement score from first-person view observations.

Motivated by the possibility of using the large number of daily interactions in the long-term deployment and exploiting the intuitive human assessments of the users' engagement during such interactions, this project proposes to:

> **Objective 3**: Devise a model of users' engagement with the technology, trained from human engagement annotations, that can be used as a proxy for automatically evaluating the robot behaviour as a measure of the quality of the interactions it delivers to the users.

The final component, closing the loop between users' engagement perception and robot behaviour, is an adaptive policy that selects the best actions that Lindsey should perform based on the state of the users. Ideally, the robot behaviour should aim to improve the interactions' quality, encouraging longer tours and eliciting high user engagement. The last objective of this project is then:

> **Objective 4**: Implementation of a Reinforcement Learning (RL) algorithm that optimises the "robotised" guided tours policy by maximising the users' engagement over the entire duration of the interaction.

By fulfilling the above objective, in addition to providing a practical computational approach for the adaptation of the robot tours, I attempt to test the following research hypotheses:

Hypothesis 3: The engagement level displayed by users indicates their willingness to stay in the interaction.
Hypothesis 4: By taking into account such users engagement, a robot can improve the quality of the interactions.

The overarching scientific contribution of this thesis is to improve the state of the art of behavioural adaptation in long-term scenarios for social robots toward having, one day, robots capable of seamlessly operating among us and learning from experience. With the presented museum guide robot scenario, I aim to validate an implementation of the proposed general framework as a proof of concept for the benefit of the whole community of Human Robot Interaction (HRI).

1.2 Contributions

This project produced several contributions to the field of HRI, as listed below.

(C1) It presented a long-term deployment spanning years of an autonomous social robot in a public environment. This is the longest-running experiment of its kind to the best of our knowledge. The long-term deployment was made possible by integrating various methodologies, monitoring interfaces and recovery procedures purposely designed to allow robust autonomous operations and minimal expert intervention. A new open-source ROS package[3] for real-time monitoring of the state of the robot, that is now being used daily in research and commercial robot deployments, was developed within this project. The deployment setup with the analysis of long-term autonomy metrics is presented in Chapter 3 and was published in **(P1)**.

(C2) It performed an analysis of users' engagement patterns with a social robot over the long-term scenario. The study's results provided a better understanding of the museum visitors' behaviour and allowed to model their engagement with the technology. The results of this analysis are reported in Chapter 3 and were published in **(P1)**.

(C3) A new dataset of human-robot interactions has been collected at The Collection museum during the guided tours provided by Lindsey the robot. The dataset is composed by RGB-D videos collected from 2 separate cameras for the entire duration of the interactions associated with a per-frame annotation of scalar engagement performed by three independent coders. The data collected is unique, considering that it was collected during unstructured interactions in a public space, and it reports a continuous engagement value intuitively estimated by coders, the first of its kind to the best of our knowledge. Chapter 4 describes how the data was collected and annotated, providing an analysis

[3] https://github.com/LCAS/sentor

of statistical significance for the annotations. These contributions were also reported in **(P2)**.

(C4) Proposes a new regression model for real-time estimation of the overall users' engagement during in-the-wild deployments from the robot point of view. The trained model is released open-source as a ready-to-use ROS node and can be used on any technology (robot, computers) where an ego-centric estimation of engagement of users is required. The model is presented and evaluated in Chapter 4 and in **(P2)**.

(C5) It proposed and implemented a robotics framework which enables autonomous social robots to improve their interactions with people online through adaptation of their behaviour. The framework is composed of a model of engagement which estimates the users' engagement state, and an *Optimistic RL algorithm* which learns the best actions to execute to maximise the future users' engagement. The overall general framework for adaptation is described in Chapter 2 and proposed in **(P3)**; the specific implementation, its evaluation are described in Chapter 5 and in **(P4)**.

1.3 Publications

Parts of this dissertation have been published in the following international journals and conference proceedings:

(P1) Del Duchetto, F., Baxter, P.E., & Hanheide, M. (2019). Lindsey the Tour Guide Robot - Usage Patterns in a Museum Long-Term Deployment. *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1-8.

(P2) Del Duchetto, F., Baxter, P.E., & Hanheide, M. (2020). Are You Still With Me? Continuous Engagement Assessment From a Robot's Point of View. *Frontiers in Robotics and AI*, 7.

(P3) Del Duchetto, F., Baxter, P.E., & Hanheide, M. (2020). Automatic Assessment and Learning of Robot Social Abilities. *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*.

(P4) Del Duchetto, F. & Hanheide, M. (2022). Learning on the Job: Long-Term Behavioural Adaptation in Human-Robot Interactions. *IEEE Robotics and Automation Letters*, 7, 6934-6941.

The following are other publications of works developed in relation to the project of this thesis that are not part of this thesis.

- Del Duchetto, F., Kucukyilmaz, A., Iocchi, L., & Hanheide, M. (2018). Do Not Make the Same Mistakes Again and Again: Learning Local Recovery Policies for Navigation From Human Demonstrations. *IEEE Robotics and Automation Letters*, 3, 4084-4091.

  This work presents an interactive framework for learning from demonstrations to detect navigation failure and generate trajectories for recovering from them. With this approach, the robot actively asks users to provide a demonstration in case of failure and then uses Gaussian Process to learn a regression model for recovering in future failure cases. This work follows the principle of enabling robots to exploit the availability of humans in long-term interactions in public spaces to improve their behaviour over time.

- Baxter P., Del Duchetto F., & Hanheide M. (2020) Engaging Learners in Dialogue Interactivity Development for Mobile Robots. In: Moro M., Alimisis D., Iocchi L. (eds) *Educational Robotics in the Context of the Maker Movement. Edurobotics 2018.* Advances in Intelligent Systems and Computing, vol 946. Springer, Cham.

  The work presents a system developed with the goal of engaging students with the design of dialogue interactions for a museum guide robot. The students were encouraged to program the high-level behaviour of the robot using Google's DialogFlow interface, which was then connected to generate behaviours on both a real robot and a simulated instance of it. Results from self-reported questionnaires show that the activity was engaging and enjoyable for the learners and that the system performed well in terms of usability. The behaviours programmed by the students were also helpful in informing the deployment of the actual museum guide robot, as described in the rest of this thesis.

- Brown, O., Roberts-Elliott, L., Del Duchetto, F., Hanheide, M., & Baxter, P.E. (2020). Abstract Visual Programming of Social Robots for Novice Users. *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction.*

  This paper proposes a high-level visual programming system for novice users, based on Blockly, for designing the robot behaviours of a museum tour guide robot, allowing to incorporate low-level social cues. Users could directly experience the results of their behaviour by executing it on the robot in the museum. Results showed that users felt the programmed interaction was more personal than the original robot behaviour and that the social cues made the robot more interesting. This study also informed the design of Lindsey's behaviour in our long-term deployment.

CHAPTER 2

Concept

THIS chapter describes the general ideas behind this thesis' work, outlining the background work in the field of robot autonomy, particularly for the long-term social scenarios. The current challenges and possibilities in such environments are described as they motivate the choices made for this project.

2.1 Background & Motivation

In recent years, various successes in robot navigation, perception and planning enabled the long-term deployments of robot systems in a somewhat autonomous fashion. Robots are more and more able to navigate dynamic environments while perceiving their surroundings and the other entities in them.

2.1.1 Long-Term Autonomy

However, even the longest running deployments with the state-of-the-art robot systems show that unexpected situations and failures are virtually impossible to avoid [40]. In such situations, there is frequent need of expert interventions which needs to either actively reprogram the robot systems to take into account the newly encountered failure situations, add and replace hardware or sensors to increase the robot capabilities of perception and computing power, or eventually fix a bug that caused the issue in the first place. In a recent tour guide robot system, similar to the one presented here, Wang et al. [95] reports an extensive amount of hardware, software and interaction failures over their one-month deployment, identifying the reliability of the robot system as one of the most crucial aspect for service robots in long-term deployment. In many cases, the unexpected situations can be resolved by non-expert humans if the robot is able to detect the issue and instruct them how to solve it. For example, when a robot is unable to navigate because its wheels are stuck on a carpet it could ask for help to a human in the surroundings for being pushed away from the obstacle, or, when a door the robot is supposed to pass through to is closed it can ask a human to open it. In the first example, the failure could also be resolved by a recovery procedure programmed in the robot (backtracking, for example) which a robot programmer can add to the repertoire of behaviours. However, in the latter the problem is unsolvable if the robot does not have the physical means to push a button and the only way to overcome it is by asking for the humans' help. This underlines 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 [76].

Overall, to ensure LTA, previous experience seems to suggest that a combination of robust software components, monitoring interfaces and reliance on the human help seems to work best.

2.1.2 Social Robots in Public Environments

When also considering those tasks where the robot is purposely meant to interact with humans, like performing a guided tour in a museum or giving directions to customers in a shopping centre, they discover that there is another entire class of failures that can happen as a result of the inability to interact socially. There are numerous factors influencing the interactions and managing the effects of them all is still a challenge to overcome. A few notable examples in the literature are listed here.

Humans' perception of a robot suffers from the "uncanny valley" effect described by Mori [67], which suggests that the users' affinity for a robot does not always increase as one makes the robot more human-like. There is, in fact, an inflexion point after which it steeply decreases before going up again, when there is a little distinction from a real human. Moreover, robots that appear similar to humans create high expectations in users about their capabilities and, whenever those expectations are not met, the effectiveness of the interaction is usually limited [38]. The interaction with the robot from the user's perspective must be intuitive and self-explanatory, whether it is mediated through speech interactions or via a display. In their long-term study, Severinson-Eklundh et al. [84] report that situations were encountered where users wanted (or needed) to interact with the robot but were unable to because they failed to understand how. Castellano et al. [22] report that current technology lacks *social perceptive abilities* – which includes "recognising people's social affective expressions and states, understanding their intentions, and accounting for the context of the situation" – in order to appropriately taking into account the unexpected conditions that occur in real-world environments.

Taken together, previous works show that in order to enable successful social interactions with humans, robots need to be designed appropriately, both in their appearance and their behaviours, so that they don't show more competence than they have. Also, the interaction needs to be self-explanatory for the user and the robot need to perceive the user "social" state in order to behave appropriately.

2.1.3 Behavioural Adaptation for Social Robots

A robot with the ability to adapt its own initial behaviour, which may not be best suited for socially interacting with humans, can learn how to improve over time given enough interactions. Behavioural adaptation is also needed because people's preferences differ for each individual and change over time. Therefore, there is no optimal behaviour that a robot can find to interact with any user, but rather a spectrum of these where each is best suited for a specific set of human preferences and behaviours. A survey on long-term interactions in social environments [55] reports that "deploying social robots in public environments, where they can interact with almost every type of person, requires additional efforts in terms of usability and adaptation". More recently, Nocentini et al. [68] identify perception and learning as two fundamental aspects that need to be addressed in future works to allow robots to perceive and adapt in real-time during interactions. They highlight the importance of conducting in-field studies where the robot system can be tested with users in dynamic environments.

This thesis focuses on robot adaptation to human preferences and behaviours while considering the users as a whole rather than personalising to the specific persons encountered, which requires (re-)identification. In public spaces, like a museum, people in the environment are continuously changing, and rarely the robot encounters the same person twice, differently from scenarios such as education, rehabilitation and personal assistance, where the robots are instead required to adapt to a specific individual.

Adapting robot behaviours to deliver better social interactions requires a metric of the quality of such behaviour in the first place in order to guide the adaptation. Here a distinction can be made between those works which use task-specific events as proxies for the goodness of an interaction, like successfully performing a handshake with the user [73] or maximising a group formation score [35], and those that are based on the perceived users' engagement or emotional state, like facial expressions [92], or joint attention [25]. Detecting task-related events from sensors is usually easier than perceiving the users' state in real-time; however, the feedback they provide is very sparse, making it difficult to evaluate the robot's actions, particularly in prolonged interactions.

2.2 Approach Overview

In this thesis, the overarching goal is to deploy a social robot to perform autonomously guided tours for the visitors of The Collection museum. Taking inspiration from the literature, a novel and unique approach has been proposed to allow the long-term autonomy and adaptation of a social robot deployed in such a public space. The general principle that informed the proposed conceptual framework is that robots should use as much as possible the users' feedback that can be obtained during the interactions –whether explicit (pushing the robot away from an obstacle) or implicit (observing how they engage in an interaction).

Despite being a challenge for ensuring autonomy, the long-term scenario is a necessary feature of the proposed deployment because it allows the robot to collect many interactions with users, enabling it to learn over time. Therefore, the first requirement for this work is to ensure LTA, allowing Lindsey to operate in the museum for several months with minimal interruption and assistance – i.e. fulfilling **Objective 1**. The autonomy of Lindsey was enabled in the first place by the development of various monitoring interfaces which were available to both the roboticists –who had access to detailed information about the robot's status and logs– and to the museum staff –who could observe the high-level status of the robot and perform few actions like sending the robot to the charging station. The monitoring interface was essential for the roboticists to quickly discover errors and failures, particularly during the first months of deployment. Additionally, the robot behaviour was programmed to take advantage of the presence of users in the environment to address many of the failures that some physical actions performed by users can solve. For example, when the robot is stuck near an obstacle, it now asks people to be pushed away from it. During such interactions, the robot technology's ability to effectively relay the information to the user in a way that is easy to understand is crucial. For this reason, special attention was dedicated to the interaction modalities, with a few examples of such features taken into account being: every speech from the robot is replicated in writing on the touch-screen; when the robot is about to move it informs the users; when the user's speech is requested, a special icon on the screen signals that the robot is actively listening.

A primary requirement of the project is that the robot should be able to engage with the users effectively. However, motivated by previous work, it is anticipated (in **Hypothesis 1**) that the users' engagement with the technology quickly fades after the initial higher attention, caused by the novelty effect, if the robot behaviour is too static. This hypothesis will be validated by enacting **Objective 2**. With this prerequisite in mind and the objective of enabling the robot to improve its ability to engage with people over time from daily interactions, it is necessary to define a metric for user engagement to evaluate the robot's behaviour.

In many works, the users' evaluations of interactions are obtained through questionnaires, handed out either by human interviewers or automatically by the robot at the end of the interactions. However, this modality is not suitable for spontaneous interactions (like the ones in the presented deployment) because many users would not agree to be interviewed, and, as a consequence, the feedback obtained would be biased by the opinions of the respondents, who were most probably the ones highly engaged in the first place. Another way of obtaining feedback from users is paying attention to certain events and user behaviours, which are used as proxies for evaluating user engagement (the user stopping a tour would indicate low engagement, for example). However, while such indirect assessment would not suffer from

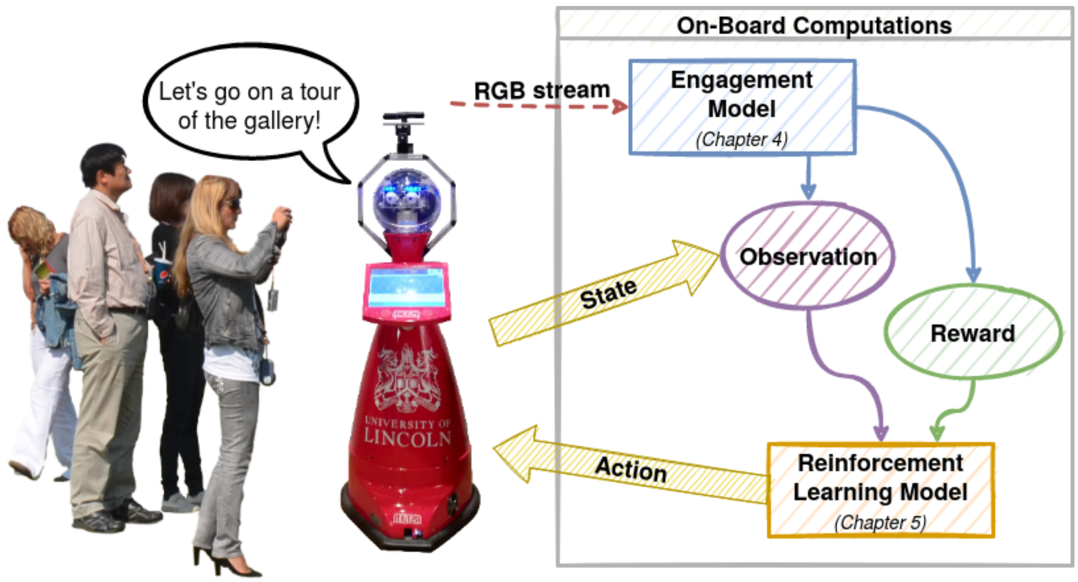


Figure 2.1: A sketch of the general framework developed in this thesis for adapting the robot behaviour from the users' feedback – in the form of their engagement – during in-the-wild interactions.[1]

the same bias of direct user interviews, it produces too sparse feedback for learning over an interaction like the tour guide, which lasts minutes. In this thesis, by attempting to verify **Hypothesis 2**, the assessment of users is established from the intuitive and holistic evaluation of the users' engagement provided by independent coders. A model of engagement is trained from such evaluations as ground truth to allow the robot to generate feedback for its own behaviour based on the users' engagement. Such measure is detected in real-time and from the robot's perspective from its camera feed so that for every action it performs, it receives a direct assessment of how it affected the users. The engagement model provides a solution for **Objective 3**. With the overall aim of developing a fully autonomous robot that incorporates human feedback during the interactions and learns from it to improve its competence, the last step of the thesis is using the engagement estimations as a reward given to the robot for its actions. Figure 2.1 shows a high-level sketch of the proposed closed-loop framework. As required by the **Objective 4**, a RL model is tasked to match state observations – of the robot, environment and the users state – with actions that are expected to elicit high user engagement in the future. This is obtained by maximising the reward received from the engagement model. By maximising such a metric, it is expected that the robot tours will be more engaging for the users and, consequently, more extended in time (validating the **Hypothesis 4**). Additionally, to test **Hypothesis 3**, the detected engagement state is included in the robot observations to condition the policy action selection on the current state of the engagement. Intuitively, different engagement levels should lead users to prefer

[1] People Pic PNG Transparent Background, Free Download #44610 - FreeIconsPNG. (2022, March 26). Retrieved from https://www.freeiconspng.com/img/44610

following a guided tour differently, with scarcely engaged people wanting to have a quick overview of the items displayed while highly engaged ones preferring more detailed descriptions.

2.2.1 The Adaptive Guided Tour Scenario

Lindsey, the robot, is embedded in an “in-the-wild” scenario where its purpose is to deliver guided tours to people in an archaeological gallery. In this setting, the museum visitors can start a guided tour from a selection of possible tours, follow the robot through the various exhibits in the tour and listen to the descriptions of the items displayed. People can also disengage and stop the tour (or simply go away) at any moment. A topological map has been defined on the museum map containing a node for each item in the robot tours. Thanks to this topology, the robot knows the location of each item in the physical world and can plan navigation trajectories to guide visitors to their locations. The items in each tour, their content and their positions, remain unchanged during the entire deployment. The robot, however, can travel to their locations and describe them in different orders and with varying levels of detail.

Given the context of interactions, in this work, the states and actions space has been limited to that of the guided tours for the purpose of learning and adaptation from the human feedback. Lindsey, the robot, is roaming around the museum, greeting people when approached and performing other tasks during idle times, however, those situations are not considered part of the learning framework in order to circumscribe the problem and study how the robot can adapt its behaviour learning from the human preferences during guide tours. Despite this simplification, the partial observability of the environment from the robot perspective is such that the planning problem is still an issue to be overcome, particularly considering the unknown behaviour of humans during the interactions.

CHAPTER 3

Lindsey & the Long-Term Deployment

THIS chapter outlines the settings under which the project has been carried out. It provides a description of the museum scenario in which the robot has interacted autonomously with thousands of people for a total period of over a year and the robot's technical details. Finally, it shows an analysis of the long-term deployment focusing in particular on the robot autonomy data and the interactions with the museum visitors.

3.1 Introduction

One of the aims of this project is the long-term deployment of a tour guide robot in a public museum. The robot is of service to visitors at all times in an archaeological gallery where it is tasked to autonomously provide guided tours and describe the items in the exhibits.

Considering the long-term aspect of the project, a certain level of autonomy is required to allow the robot to be available at all times during the museum's opening hours. In this context, remote teleoperation techniques, like Wizard-of-Oz [49], are not feasible since they would require a person to continuously control the robot. However, the presence of people (which is a requirement for robotics application in social contexts) in such a public space is a source of uncertainty to the robot's perception of the environment, but also in the way the interactions with the users unfold. The interactions between the humans and the robot are in fact, spontaneous –i.e., users are not instructed how to behave with the robot and are allowed to freely approach the robot, initiate and pursue interactions, or leave at any time. A range of mechanisms, described later in this chapter, have been developed in this project to allow the robot to be robust in such an uncertain scenario. Moreover, the engagement of users with the robot is analysed to evaluate not only the robot's functionality but also it's effectiveness at interacting with users.

3.2 Related Work

In the past, several works have addressed the problem of Long-Term Autonomy (LTA) in robotics demonstrating the extent to which robotics frameworks can enable robots deployments for long periods of times in various scenarios.

Luperto et al. [61] present a study to evaluate a long-term deployment of an autonomous system in an apartment for assisting elderly users. Altough their work is only a preliminary experiment preceding the actual 3-months deployment in the users' houses, it provides a detailed breakdown of the system developed which includes the coordination of multiple behaviours the robot must perform, such as actions triggered by a *Virtual Caregiver* AI, by the users or self-management actions triggered by the robot itself (for example, docking to the charging station). Meeussen et al. [64] have deployed a robot for 13 continuous days in an office environment while exploring ways to improve the robot's robustness by identifying failures and recoveries (including asking people's help). In [15], a fleet of four CoBots reached 1,000 km of overall autonomous navigation. The robots were able to seek human assistance to perform manipulation tasks (the robots did not have arms) and send emails to developers in case of a lack of human response. Within the STRANDS project [40], the SCITOS G5 robot travelled more than 160 km over three individual deployments. Although the robot was able to be autonomous most of the time, the authors report the need for a way to manage failures and to have a better understanding of human activities. These works have demonstrated that long-term autonomous deployments, although potentially viable with current technologies, still pose challenges to overcome. In particular, with the management of failures remaining an open problem it seems that the harnessing of users' help in such situations is the most common recovery. In the presented deployment, taking inspiration from these works, the human recovery is in fact integrated in most failure management strategies where the resolution does not require expert intervention.

The long-term aspect enables to optimise the chances of encountering humans during operations. For example, Hanheide et al. [39] propose a spatio-temporal model to learn when, where and how users interacted with the robot info-terminal during a long-term deployment. They found they could improve the efficiency and usefulness of the system by proposing the right content at the right time and place. To solve failures, Rosenthal et al. [77] presents a model that exploits the availability of humans in the environment by planning path that increases the likelihood of encountering of an active helper. The presence of humans, although helpful for failure recovery, can pose challenges to the robot autonomy as well. For example, Luperto et al. [61] found that people in order to interact with their robot positioned themselves in a challenging position for the navigation algorithm to be able to move without failures. Although they report that the robot was able to successfully recover from all those failure cases by simply replanning a new path, the limited duration of the study (9 days) and the fact that the robot was tested only in one small environment does not allow to generalise the results for long-term deployments and more complex deployments. Especially when robots are tasked to interact socially with humans, focussing only on autonomy and functionality is not enough; the ability of the robot to engage and interact appropriately with the users is of high importance as well. A survey on long-term interaction between users and robots [55] raises the issue that memory and adaptation for social tasks remain nearly unexplored in the field. Similarly, Kunze et al. [52] in exploring the state of the art in Artificial Intelligence (AI) techniques for LTA, asserting that a significant challenge for future deployments is integrating human interactions in the robot system to allow improving its knowledge in future situations. Wrongly managing the users' expectations is a first barrier to ensuring the robot social interaction is effective toward non-exper users. Gockley et al. [38] have deployed a robot receptionist for a long period of time, discovering that although a few visitors repeatedly engage with the robot, most only interacted until the range of capability exhibited by the robot is exhausted indicating that they initial expectations of the robot where higher. They propose that robots should be endowed with personality and character, and be able to personalise to the user to keep them engaged for longer. The way the interaction is carried out with the user also must be accessible and self-explanatory. Severinson-Eklundh et al. [84] found that some users were not able to start an interaction with their robot in a long-term deployment because they failed to understand how. Explainability is crucial in interactions also when the robot seeks the users' help; however, cognitive, psychological and social determinants that impact the design of communication and failure mitigation strategies are not well-studied [44]. Eventually, to enable natural and engaging human-robot interactions robots need to possess the ability to perceive the user's affective state, their intentions and account for the context of the interaction; however, these capabilities are still lacking in current social robots [22]. In this thesis, a robotics system is presented that accounts for the online assessment of the users' social state (in the form of their engagement) and that adapts the robot actions accordingly to increase the amount of time users decide to stay in the interaction. This system is based on the framework designed in this chapter and supported by the initial report on the users' engagement with the tour-guide robot before the adoption of an engagement assessment and adaptation framework.

Looking specifically at deployments in museum environments, many examples of long-term deployment have been found. The robot Rhino [18] was deployed in a museum in Germany for six days guiding hundreds of visitors. At the time, the main issues and the focus of the work were navigation and obstacle avoidance. The Minerva robot [93] traversed more than 44 km and interacted with more than 50k people. Moreover, it was able to display mood (i.e. happy or angry) and used an RL approach to learn the best actions to engage visitors. In [69], four robots were deployed over five years. Focusing on interactivity and education, they learned that: 1) there is usually a crowd around the robot, therefore, any speech from the robot should be multimodal; 2) long and non-interactive presentations are guaranteed to

Figure 3.1: Pictures of the archaeological gallery.[1]

drive the audience away. By taking inspiration from previous lessons, this work presented a robotic system that enabled accessibility to the users during the guided tours and robustness in navigation, utilising the users' help. Notably, Webster et al. [96] explore the perception of the public on the appropriateness of using robots in museums and galleries. They report an essential finding: people find using robots in such environments more appropriate when they expect the robot to have emotional skills. This suggests that service robots in museums need not only to be able to be autonomous but also to understand the users' social state and behave accordingly, possibly expressing their own emotions. In the present work, some of these challenges are addressed by enabling the robot with the ability to directly optimise the users' engagement during the interaction by allowing the robot to adapt its own behaviour to the users.

3.3 The Museum Scenario

The environment used for the works in this thesis is the archaeological gallery of The Collection museum[2] in Lincoln, UK. This permanent section of the museum, shown in Figure 3.1, is freely available to any member of the public remaining open 7 days a week (6 days a week after COVID lockdown reopening) from 10 AM to 4 PM.

The gallery hosts thousands of local archaeological findings dating from the Stone Ages to the Early Modern Period. Lindsey the robot can describe and guide people to a subset of them which includes items ranging from different periods and categories. These items are shown grouped per category in Figure 3.2 and their position in the gallery is shown in Figure 3.3.

The environment presents a considerable number of challenges for the robot autonomy both in terms of its ability of remaining functional and available for interaction and in its capability of interacting socially with spontaneous users in such an unstructured environment. Visitors of the museum can freely explore the gallery and are not informed in advance about the robot presence in the environment. For most

[1]View our museum in 360° | The Collection. (2022, April 05). Retrieved from https://www.thecollectionmuseum.com/hire-us/photography-360

[2]https://www.thecollectionmuseum.com/robot-at-the-collection

Figure 3.2: The items in the gallery present in the robot's knowledge base.

visitors, who do not know that a robot is in the museum, the encounter with the robot is entirely spontaneous. People are also not instructed about how to behave with the robot or what its functionalities are. They can freely start an interaction with Lindsey, in the same way, they can just leave the robot-guided tours at any moment if they want to.

3.4 The Lindsey Interactive Mobile Platform

3.4.1 The Robot System

The robot used in this project is a Scitos G5 model produced by MetraLabs GmbH[3], called Lindsey. Figure 3.4 shows the robot during the deployment at the museum. The platform has a mobile base which allows to navigate the environment and its robust structure can operate for various hours and kilometers autonomously. The version used in this project is equipped with a touchscreen, cameras, microphones, speakers and "human-like" eyes to interact with the users. It is CE certified for

[3]https://www.metralabs.com/en/mobile-robot-scitos-g5/

Figure 3.3: A simulated model of the gallery with markers indicating the positions of items visited in the robotic tours.

autonomous and safe navigation. The mobile base is non-holonomic and can reach a speed of 1.2 m/s. Its battery allows about 8 hours of autonomy.

In order to ensure safety during navigation the robot base is protected by a rubber band which cuts the power to the motors when detecting a pressure. Similarly, the robot body has two easily accessible emergency buttons that, when pressed, stops the motors. For navigation and localisation the robot uses a laser scanner with a 270° scan angle. Additionally, the perception of obstacles and depths in the floor is facilitated by an Intel Realsense depth camera positioned in the front of the robot base. Two additional depth cameras, positioned on the robot head and above the touchscreen, detect people during the interactions. To support the computational needs for autonomous operations the robot is equipped with 2 computers: the central unit, which communicates with the robot electronics and executes the main software components (such as task scheduling, navigation, and monitoring) and, a secondary unit which is equipped with a GPU and supports the central unit with running accessory processes (such as face detection). The software system is based on the Robot Operating System (ROS) framework and runs on the Ubuntu 16.04 Operating System. Many of the ROS nodes in use are based on the STRANDS project [40] core modules. In particular, the system uses the STRANDS packages for logging data of the operations into a Database, Task Scheduling and the robot roaming behaviour.

3.4.2 Robot Behaviours Specification

The operations that the robot performs are specified in a hierarchical manner, where at the highest level there are *tasks* which consist of self-contained operations that are extended in time, such as a "performing a tour guide" or "go to charge the

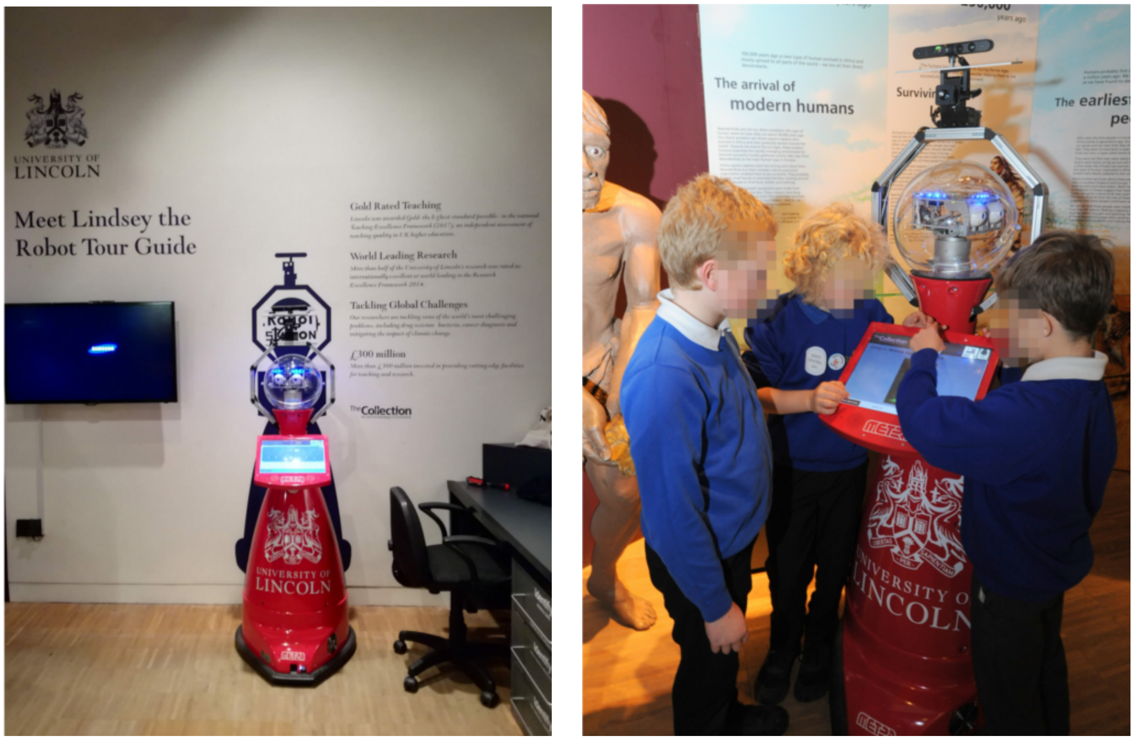


Figure 3.4: (left)Lindsey at its base station. (right)Students interacting with Lindsey.

battery". Each task has an associated priority, expected duration and a time window for execution. Whenever a task is demanded, a scheduler takes care of deciding when and if it should be executed, taking into consideration the other tasks present in the schedule. A task can take precedence over another that is already scheduled if it has a higher priority.

The tasks are a composition of lower level behavioural units and are defined as *conditional plans*. The lower level units of conditional plans can be other (sub-)plans or actions –which represent the lowest level of atomic behaviour the robot can perform.In order to enable such compositionality of the robot behaviour the plans have been specified using the Petri Net Plans (PNP) formalism.

Petri Net Plans

Conditional plans are defined using the Petri Net Plans (PNP) formalism which was chosen for being well suited to represent robot behaviours, offering a rich set of features such as concurrency, interrupts and sensing operations [100]. PNPs can be easily translated into stochastic processes allowing learning and adaptation of the robot behaviour [59]. In the museum deployment an implementation of the PNP formalism which was adapted, by the author of this thesis, was utilised to allow better compositionality and reusability of the plan components and for improved integration with the ROS middle-ware.

The formal definition of the language is described in [100]; for the purpose of this

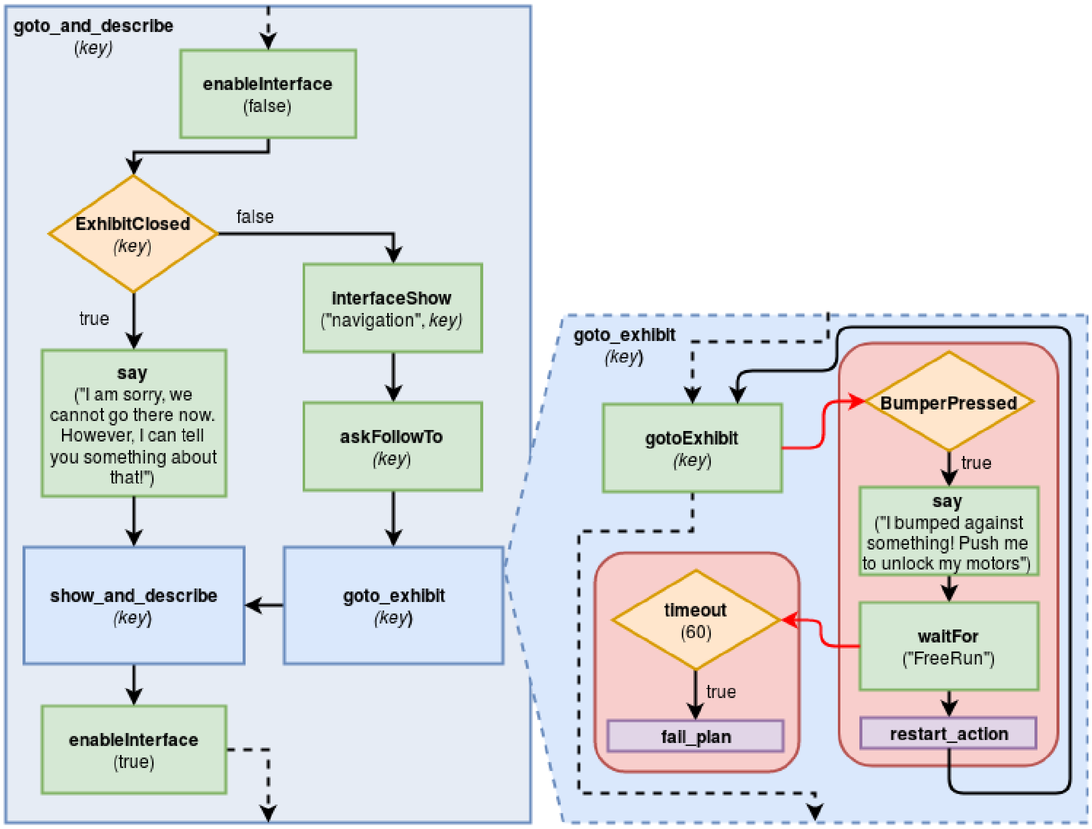


Figure 3.5: Block diagram to represent visually the plan `goto_and_describe`, with the expansion of the sub-plan `goto_exhibit`. Green nodes are low level actions, orange nodes are conditions, light-blue nodes are plans and purple nodes are execution rules terminal statements. Red containers enclose execution rules. Arrows shows connections between actions/sub-plans, conditions and statements. Black arrows coming from actions/sub-plans are only traveled after the source action/sub-plan is executed. Red arrows are connection that are active while the source action is in execution. Values between round brackets are the parameters passed to the action/condition/plan.

thesis the underlying mechanisms is abstracted and the focus is on the syntax only. The essential components of the plans are actions $a \in \mathcal{A}$, which corresponds to the low-level capabilities that the robot system can execute, and conditions $c \in \mathcal{C}$, which are the truth values of certain aspects of the state (i.e. the state of the environment or the robot system). A plan is then a sequence of actions and conditions. Actions, can be started, stopped and they modify the state of the environment and/or the internal state of the robot with their execution, while conditions are evaluated in boolean expressions to steer the flow direction of the actions execution. Each plan is also a sub-plan and can be included in other plans like actions. In order to interrupt the execution of actions, when certain unexpected conditions appear, the syntax allows to attach *execution rules* to actions [45]. Such rules are defined as tuples (a, ϕ, σ, ρ) where ϕ is a boolean expression, σ is an action or a sequence of actions, and $\rho \in \{$`restart_plan`, `restart_action`, `skip_action`, `fail_plan`$\}$ is a statement which determines how to continue the execution of the plan. Whenever ϕ is satisfied during the execution of a, the action is stopped and the sub-plan σ is executed. ρ determines how to resume with the plan afterwards.

Actions, conditions, execution rules and the plans themselves can be implemented in the Python[4] language using the PetriNetPlans library [5]. Actions are implemented and executed independently of each other, in fact multiple actions can be launched concurrently. In writing the plans, standard Python operators for flow control –such as "if", "for" and "while" operators– can be used in conjunction with the implemented conditions. To ease understandability, Figure 3.5 shows a block diagram representing the plan `goto_and_describe` which is used to guide people to an exhibit and, upon arrival at destination, describe the exhibit's item. The diagram shows how sub-plans can be included into larger plans, and how execution rules can be used for example to catch a sensor event during navigation.

3.4.3 Routine robot operations

Routine operations are those tasks that the robot autonomously executes when it is not interacting with people. Some of them are scheduled through a calendar in order to be executed at a specific starting time and with a particular duration, others are executed during idle times.

Every morning, when the museum opens, the robot executes the *Tour guide activation* task which gives instructions for autonomously leaving the charging station, activating the user interface and starting saving logs of the robot system. Similarly, at closing time the *Tour guide deactivation* task instructs the robot to return to the charging station, deactivate the interface and upload all the data collected during the day on the cloud server. The *Tour guide activation/deactivation* tasks are time sensitive and hence scheduled through the calendar utility.

At idle times, the robot goes around the gallery to increase its chances of meeting visitors and interacting with them. For this purpose, the *Roaming* task is scheduled to be executed every 5 minutes in different places of the gallery. The robot navigates to the given location and waits to be approached by someone; when that happens, it prompts users to start an interaction using the robot's touch-screen. The task priority is set to the lowest possible, to allow any other task to override this idle behaviour. The location where the robot is sent to at any time is decided by a spatio-temporal model, proposed by Hanheide et al. [39], which learns where the robot is most probable to start an interaction at the given moment.

[4] https://www.python.org/
[5] https://github.com/francescodelduchetto/PetriNetPlans

3.4.4 User Demanded Tasks

When users approach the robot, they can start one of 3 different interactive tasks –namely *Guided tours*, *Go to exhibit and describe* and *Describe item*– which is then scheduled for immediate execution with high priority, eventually overriding any routine task currently in execution. For the purpose of adaptation of the robot behaviour, i.e. the final objective of the project described in this thesis, only the *Guided tour* tasks will be taken into consideration in the subsequent chapters, as these allow for more possible variations to be learned being a more complex and extended in time.

Guided tour

A tour is composed of a sequence of exhibits linked by a common theme. The tours themes, the content presented in them and the order in which the items in the tours are described were defined before the deployment started with the assistance of the museum's educators. Therefore, the sequence of actions in the tour and the content in them does not change over different sessions. The tour starts with a verbal descrition of the theme chosen in order to provide context for the sequence of items shown during the interaction. After the theme description the robot guides in sequence the visitors to each exhibit of the tour. At each stop of an exhibit the robot gives a short initial description of the item shown; then it asks to the users whether they would like to have more details about it or to move to the next exhibit. The visitors can reply with the touchscreen through a yes/no modal window or by verbally pronouncing their answer. In case the robot does not receive an answer within 1 minute it terminates the tour, assuming that the visitors have left. At each exhibit, alongside the verbally description of the item, the robot shows images from its display to enhance understandability. In case that one or more stops in the tour are not acessible to the robot, the users are first guided to those that are available, and successively the others are described without physically moving to their location. In that case, the robot informs the user of the area blockage.

Go to exhibit and describe

The robot guides the visitors to an exhibit of their choice and then describes it (verbally and by showing images on the display). Like for the *Guided tour* stops description, the robot offers to provide additional information to the visitors and it does so in case they accept. If the robot is forbidden to travel to the area in which the exhibit is located, it describes the exhibit without reaching the location (informing the user of the location blockage). These tasks are considered here for the study of the visitors usage patterns of the museum, but are not core part in the following chapters.

Describe item

The robot gives a short verbal description of the exhibit demanded by the visitor. These tasks are considered here for the study of the visitors usage patterns of the museum, but are not core part in the following chapters.

3.4.5 User interface

The main interaction modality between the robot and the users is the touchscreen mounted on the robot body. The Graphical User Interface (GUI) displayed on the screen is a web application implemented in HTML/CSS and JavaScript that communicates with ROS through the roslibjs library[6]. From the GUI, users can browse the different robot tours and visualise an interactive map of the museum with the position of the items that the robot can describe. Figure 3.6 shows the three pages the users can navigate through for starting an interaction.

After the user has visualised what tasks the robot can perform, they can decide to start one of the user-demanded tasks described in the previous section –namely a "guided tour", a "go to exhibit and describe", or "describe item". When a user demanded task is started, the interface shows contextual information about the robot's action in execution. For example, when the robot is guiding users to the location of an item the GUI shows the text "Follow me to the *Bronze Age Barrow Finds*"(where *Bronze Age Barrow Finds* is the name of an item chosen as an example) and a picture of it; when the robot is describing an item it shows pictures that enhance the understanding of the item (rather than showing the image of the real item displayed in the exhibitor). Every time that the robot speaks, a written speech dialog is showed on the monitor to increase accessibility. Moreover, notification badges appear on the corner of the screen in the event of failure to given an explanation to users for the possible abnormal behaviour of the robot. The interface also allows users to stop the robot task currently in execution with a button that is always shown in the footer of the screen. Figure 3.7 shows two screenshots of the interface during a guided tour.

In some moments during the interaction the robot takes ownership of the interface by disabling all the touch events of the GUI components (except the "stop task" button), effectively preventing the users from browsing through the pages or starting other tasks. This design choice was made to prevent inconsistencies with the tour experience during the interaction. The users still have the possibility of starting a different interaction, but they can do so by first stopping the current task. Taking inspiration from such requirements and the lessons learned during the deployment a JavaScript library, ROS Web Components[7], was developed by Laurence Roberts-Elliott for quickly generating robot interfaces automating most of the underlying work needed for communicating with the ROS system. The ROS Web Components

[6] http://wiki.ros.org/roslibjs
[7] https://github.com/laurencejbelliott/roswebcomponents

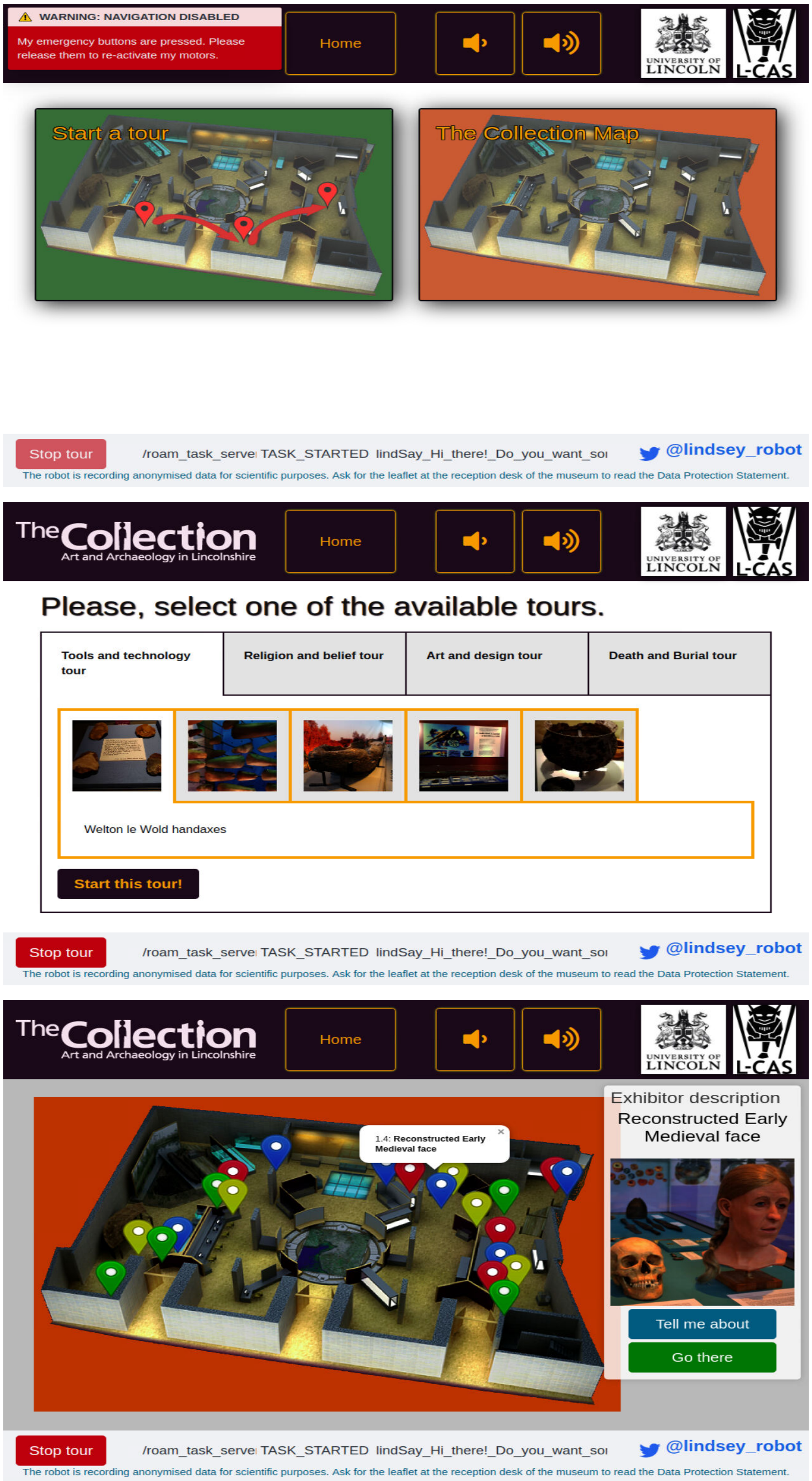


Figure 3.6: Screenshots of the graphical interface diplayed by the robot's touch screen. (top)Home page, (center)tour selection page, and (bottom)map with the items page. Warnings requiring the users' attention are presented on the screen as shown on the top image.

Figure 3.7: Screenshots of the graphical interface displayed by the robot's touch screen during a guided tour. (top)During navigation toward an item location, and (bottom)during a description of the item. Any robot utterance that comes out of the speakers is replicated on the screen, as shown on the bottom image.

library was instrumental for the development of a *Visual Programming* interface that was used in Brown et al. [16] for allowing novice users to develop robot behaviours.

3.5 Assuring Autonomy

The first step in deploying an autonomous robot in a public and dynamic environment is to ensure a certain level of autonomy so that the robot can be operational at most times and requires minimum assistance from the robotic experts. Having already established in Subsection 2.1.1 that all failure situations are impossible to prevent, in the current project the following principles have been followed to assure the best autonomy possible:

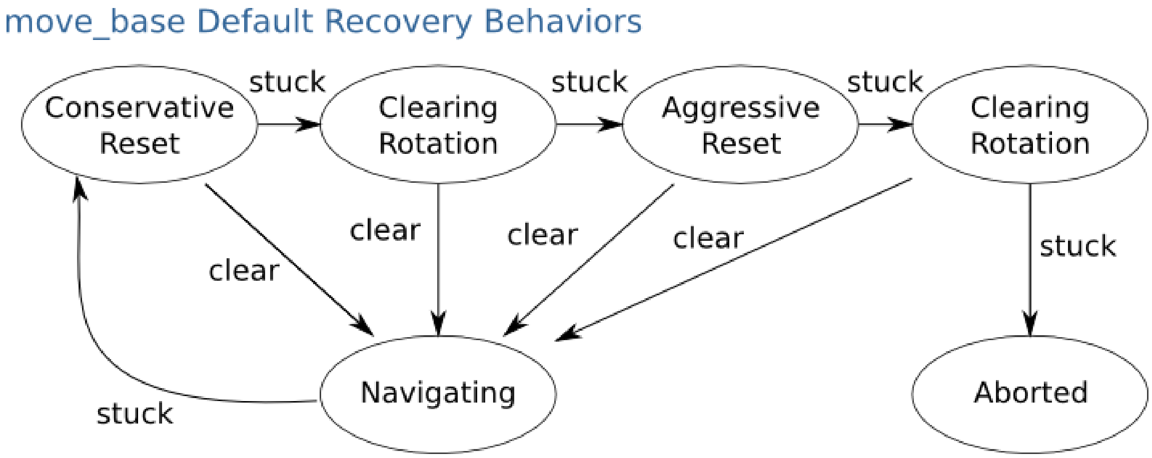


Figure 3.8: `move_base` recovery strategy for navigation failures[10].

1. design the robot's software components and its behaviours to automatically handle known exceptions with recovery strategies;

2. when the exceptions cannot be internally solved by the robot system but can be easily solved by non-expert intervention, prompt the users for help and empower them with the information needed to solve the issue;

3. as a last resort, the robot should immediately notify the relevant competent people (i.e. the roboticists and the museum's supervisors) about the issue so that it can be addressed quickly.

3.5.1 Recovery Strategies

Navigation is one of the main sources of failures for mobile robots deployed in dynamic environments. In the present work, the `topological_navigation`[8] planning framework is utilised, and `move_base`[9] is used at a lower level for metric navigation. The topological navigation package has already an internal strategy for recovery, as shown in Figure 3.8, however, this is commonly not enough for autonomous navigation in dynamic environments, as reported by Hawes et al. [40]. Failures can still happen because of sensing inaccuracies that lead the robot to hit an obstacle, people pressing the robot's emergency stops, or the planner being unable to generate a path in a tightly crowded room, for example.

In the deployment presented in this work, most errors and failures happen during navigation because 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, in this thesis also social failures are considered,

[8] https://github.com/LCAS/topological_navigation/
[9] http://wiki.ros.org/move_base
[10] move_base - ROS Wiki. (2022, April 14). Image retrieved from http://wiki.ros.org/move_base

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 addressed:

- *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. When they are pressed the robot actively asks the users to release them when a navigation action is executed.

- *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 sightly 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.

- *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 to a zone clear from obstacles, it restarts the navigation action.

- *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 causes 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.

- *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 recovery procedure for the robot consists in 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 exhibit in the tour). If the robot does not receive any response after one minute, it assumes that the users have left the tour, and hence terminates the interactive task.

The exceptions listed above are examples of situations where a planning framework that allows the design of complex robot behaviours, like PNP used in this project, provide an effortless way of monitoring, handling and recovering from such situations. For example, Figure 3.5 shows a visual representation of the sub-plan handling the *bumper pressed* using the *execution rule* constructs of the planning framework.

3.5.2 Management interface and remote monitoring

During Lindsey's daily operations, the robot engineers and the museum staff members have the possibility of remotely monitoring the robot through a web interface, shown in Figure 3.9. The home page gives a glimpse of the robot's state by showing the position on the map, the RGB camera feed, the task in execution, and a few other data. A second page of the interface allows to control some aspects of the robot operations by clearing all the scheduled tasks, immediately sending the robot to the charging station and blacklisting some areas of the gallery to prevent the robot from moving into it. The interface also shows a calendar with the robot's routine tasks scheduled in advance to be executed at a specific time and with given duration.

The web application is served over the internet from a computer located in the museum, where is can access the robot data in real-time. For this, the interface PC runs a ROS instance and communicates with the robot using the roslibjs library.

3.5.3 Critical events notification

In a long-term deployment of an autonomous robot, that operates without constant supervision, it is essential that the robot engineers are promptly informed when their intervention is required. The notification for support from the robot system needs to be both timely –so that the appropriate actions can be taken as soon as possible– and informative –so that the situation is assessed remotely and the correct action executed. To provide such functionalities in this project, a novel software module was developed: Sentor[11]. Sentor allows developers to define custom conditions based on ROS topics' messages for when a notification should be sent. Conditions can rely on a ROS topic message being published (or not published) or on custom lambda expressions using the content of the topic messages. For example, for the museum deployment, a rule was defined to send a warning every time the robot's head camera stops sending RGB images, and when the covariance of the robot localisation is above a critical level. Each notification rule in sentor can be automated with actions to be executed, calling a ROS Service for example, and with custom log messages to improve the understandability of the issue. With such a tool, the robot developers could easily define new notification rules as new exceptions were discovered during the deployment and remove outdated ones. Instead of observing the usual high-bandwidth stream of ROS errors/warnings from the system, Sentor allows developers to effectively filter out unnecessary warnings and focus on the most important ones only.

[11] https://github.com/LCAS/sentor

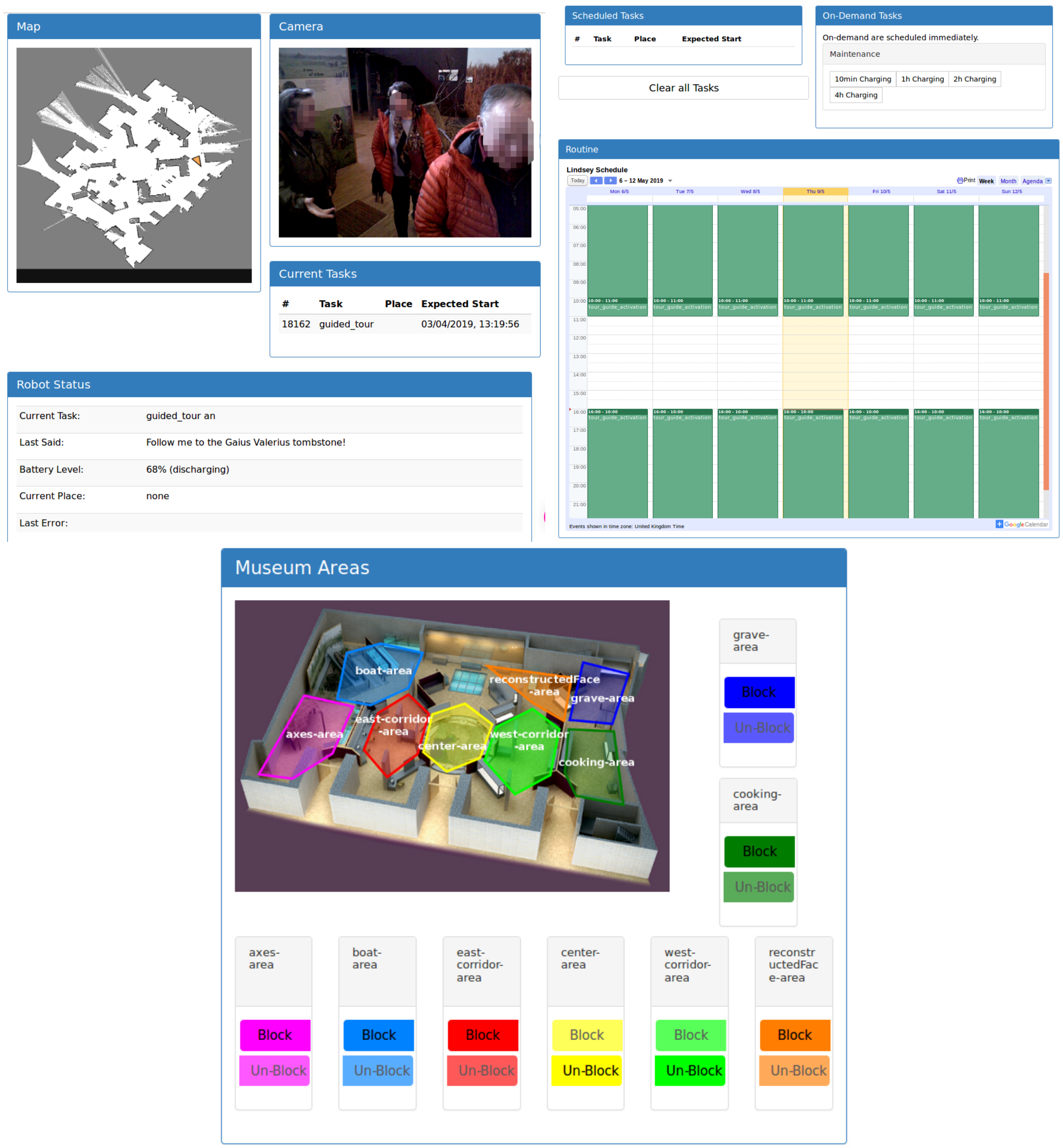


Figure 3.9: Screenshots of the management web interface. (top left) The home page shows the position of the robot, the camera live-feed and some other information about the robot's state; (top right) the tasks page allows monitoring the currently scheduled tasks and the routines that are managed by the calendar, with the possibility of clearing all the tasks scheduled or temporarily sending the robot to the charging station; (bottom) a tool for blocking areas of the museum in which the robot can not navigate.

Table 3.1: Number of user-demanded tasks with their duration. Between parentheses are reported the values obtained before filtering out the tasks in which a failure has occurred.

Task	Tot. demanded	Avg. duration	Shortest	Longest
Guided tour	5008 (7876)	3.99 (4.54)[m]	4.5 (1.43)[s]	18.9 (23.4)[m]
Go to exhibit	4632 (6652)	1.7 (1.8)[m]	7.8 (7.5)[s]	11.5 (30.8)[m]
Describe exhibit	1601 (1673)	21.2 (24)[s]	7.3 (7.3)[s]	41[s] (5.8[m])

A second software tool, namely SlackeROS[12], takes care of sending a Slack[13] message to the robot developers feeding from Sentor's notifications. Each message is accompanied by a current picture taken from the robot's head camera to give context about the robot's situation.

3.6 Deployment Analysis

This section analyses the long-term deployment of Lindsey in The Collection museum, focusing in particular on the ability of the robot to maintain autonomy, eventually leveraging human assistance, and on its interactions with the users. The purpose of such analysis is to study whether the **Objective 1** was fulfilled –i.e. a robust autonomous long-term deployment– and to execute **Objective 2** –a deployment analysis focused on the users' engagement with the robot.

The data shown in this section refers to the date range between the 24$^{\text{th}}$ January 2019 (the day on which data recording of the robot operations started) and the 17$^{\text{th}}$ May 2022. This time frame also includes some long periods of interruptions of the robot operations, as detailed in Figure 3.6.2, for which no data is reported. The robot has been operative in the museum since October 2018, however, data collection was not in place during the initial months.

3.6.1 User Interactions Performances

To evaluate the performance of our robot in engaging the users, this section analyses user-demanded tasks (see Subsection 3.4.4 for a description of the user-demanded tasks). Although directly measuring the users' engagement level with the robot in-the-wild is out-of-reach with current technologies, it is estimated by studying the robot usage, which is a component of user engagement and an easily accessible value to measure.

Table 3.1 shows, for each task class, the total number of instances started by the visitors reporting the average, shortest and longest duration. As discussed in Subsection 3.4.5, the interface allows the users to stop a task at any moment during its execution; therefore, the tasks' duration can vary between different instances.

[12] https://github.com/marc-hanheide/slackeros
[13] https://slack.com

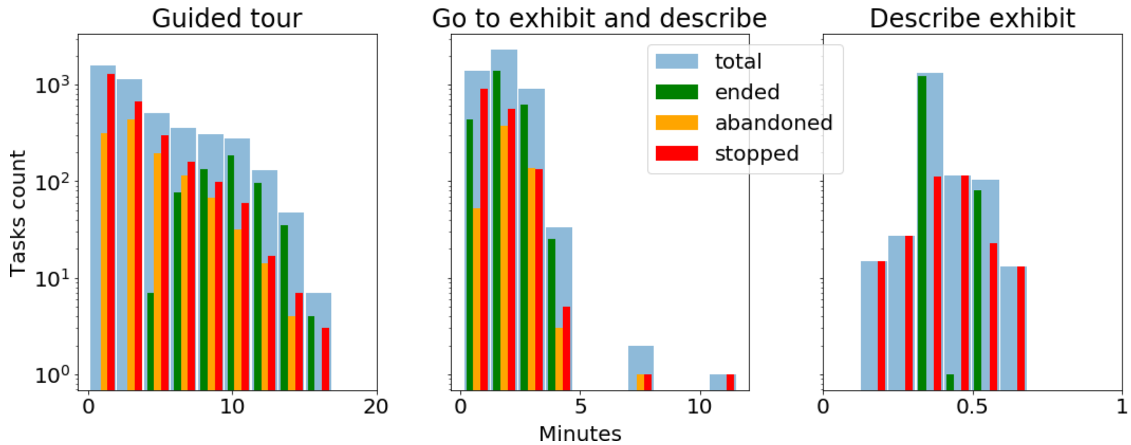


Figure 3.10: Duration distribution (in log scale) of the interactive tasks performed by the robot. The different colours green, orange and red represent the tasks that were completed to the end, abandoned by the users or stopped by the users, respectively. The light blue bars sum up the three task groups.

The table reports statistics about the entirety of the user-demanded tasks, including those that have been stopped or abandoned. A normal *Describe exhibit* task typically lasted about 20 seconds; the shortest tasks are those stopped by the visitors while a software bug caused the longer ones. *Go to exhibit and describe* tasks lasts typically about 1.8 minutes. In this case, the duration of the shortest tasks is caused by the visitors stopping them or by the fact that the robot did not navigate to the exhibit location because the area was blacklisted. The most prolonged tasks can be caused by software/hardware malfunctions or by the navigation action taking too long to reach the destination due to being affected by unpredictable obstacles and failures.

Figure 3.10 shows the duration of all the instances grouped according to how they terminated. Abandoned tasks are those stopped by the robot when it detects that the users walked away, as described in Subsection 3.5.1. A *Guided tour* task which completed to the end (not stopped nor abandoned) has a duration of about 10 minutes. The longer and more complex the task is, the higher the chances it will be stopped by the users, as can also be observed from Figure 3.11. Analysing the PNP actions tasks that were in execution when users stopped a *Guided tour* task, in Table 3.2, it is evident that tours are stopped consistently more often during the navigation actions and when the robot is describing an exhibit's item.

Analysing the tours that were executed on different days of the week, enables to discover patterns of the users visiting the museum. From Figure 3.12, it can be observed that there were far more museum visitors starting a guided tour during weekends, with Tuesday and Wednesday being the least busy days. Conversely, on average, the group of people interacting with the robot was more numerous on Tuesdays and Wednesdays than on the other days of the week. Looking at the theme of the tours started and at the specific exhibits in each tour, Figure 3.13 shows that the *Death* tour is the most popular and that the exhibits appearing first in the tours'

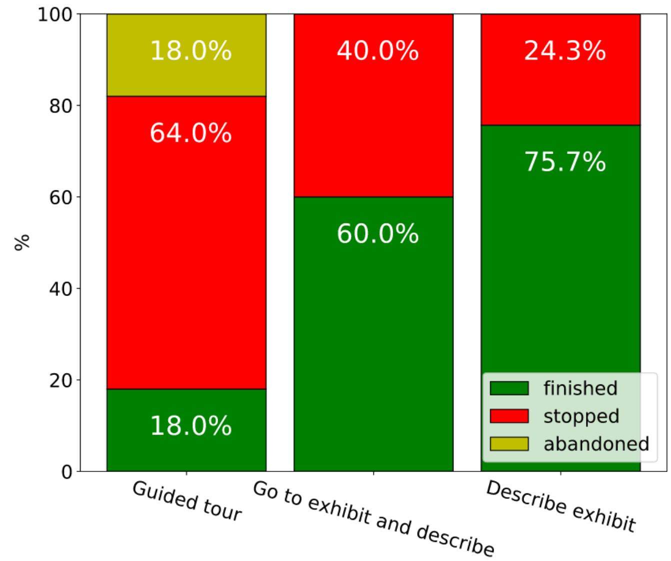


Figure 3.11: Rate of tasks normally finished, stopped, or abandoned by the visitors.

Table 3.2: Actions in execution when the user stopped the tour with associated number of occurrences.

Action	Stopped tours
GO_TO_EXHIBIT	496
DESCRIBE_EXHIBIT_ADDITIONAL	229
DESCRIBE_EXHIBIT	229
PROPOSE_ADDITIONAL_INFO	94
ASK_HELP_BUMPER_PRESSED	93
DESCRIBE_TOUR	75
ASK_FOLLOW	74
NO_ACTION	51
ASK_HELP_EM	20
ASK_HELP_NAV_FAILURE	19
TOUR_ENDING_SENTENCE	18
THANKS_HELPED	2

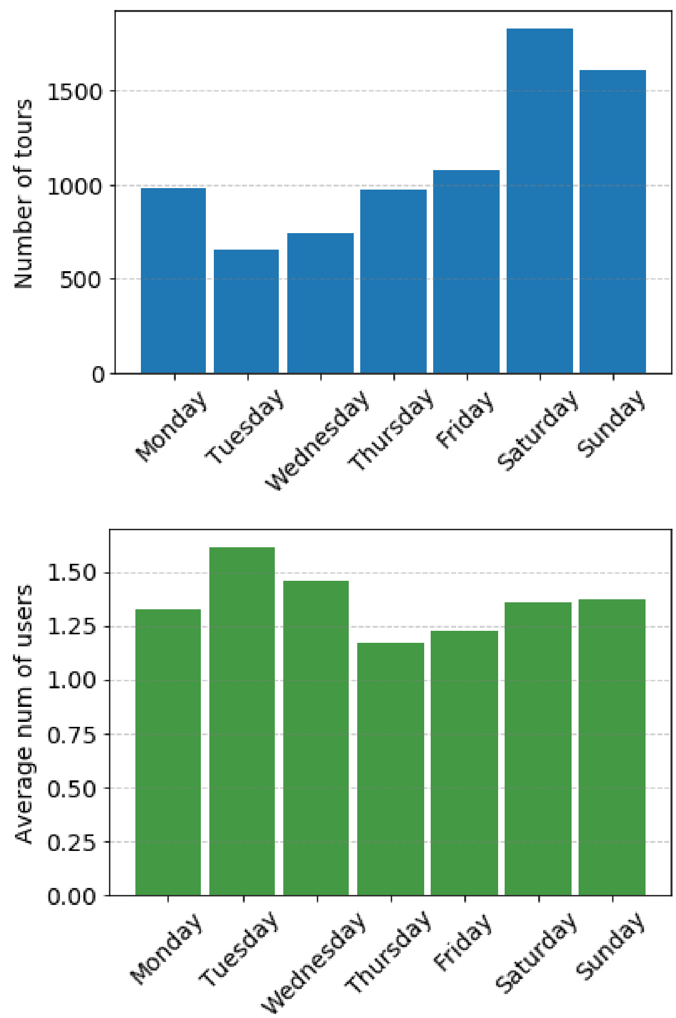


Figure 3.12: (top)Number of tours started for each week day, (bottom) average number of users during the tours per week day.

sequences are the most visited ones. This result follows the intuition that the latter items are less frequently visited because people can stop or walk away from the tours.

3.6.2 Autonomy Performances

To assess the robot's performances in terms of Long-Term Autonomy (LTA) during the current deployment, this section reports the overall system performance against two metrics: *total system lifetime* (TSL) and *autonomy percentage* (A%), as previously done in the seminal work of Hawes et al. [40]. TSL measures how long the system is available for autonomous operation, A% measures the duration the system was actively performing tasks as a proportion of the time it was allowed to operate autonomously (which in the context of this deployment is restricted to the museum opening hours). Table 3.3 shows these metrics along with other LTA measures.

Figure 3.14 shows the complete schedule of the robot's tasks executed in the first months of the deployment during the opening hours of the museum. Given that the

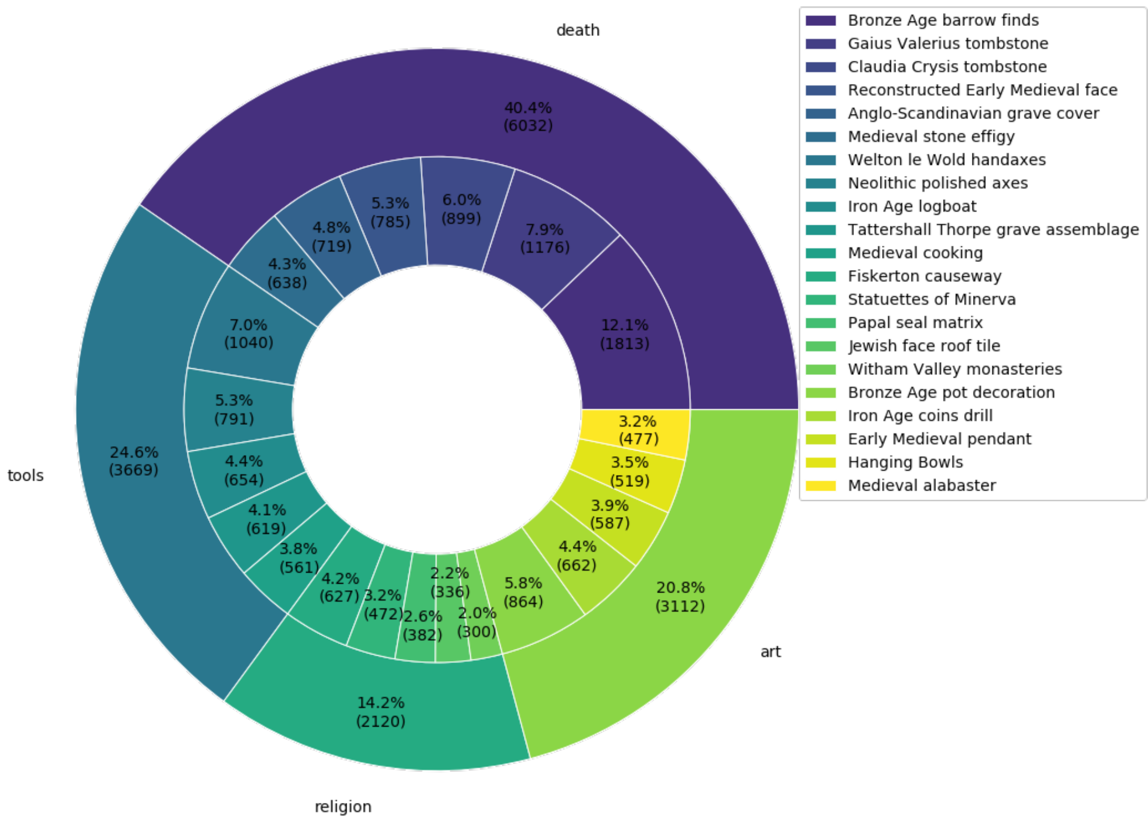


Figure 3.13: The amount of tours started for the different tour themes and the amount of items visited in each tour. Specific items are displayed in the order of appearance in the static tours, starting from darker to lighter colours.

Table 3.3: Long-Term Autonomy metrics.

Total deployment duration	1209 days (3.3 years)
Days of operation	446 days (1.2 years)
Total distance travelled	1118 km
Total tasks completed	31154
TLS	2765 hours (111.46 days)
A%	75.8%

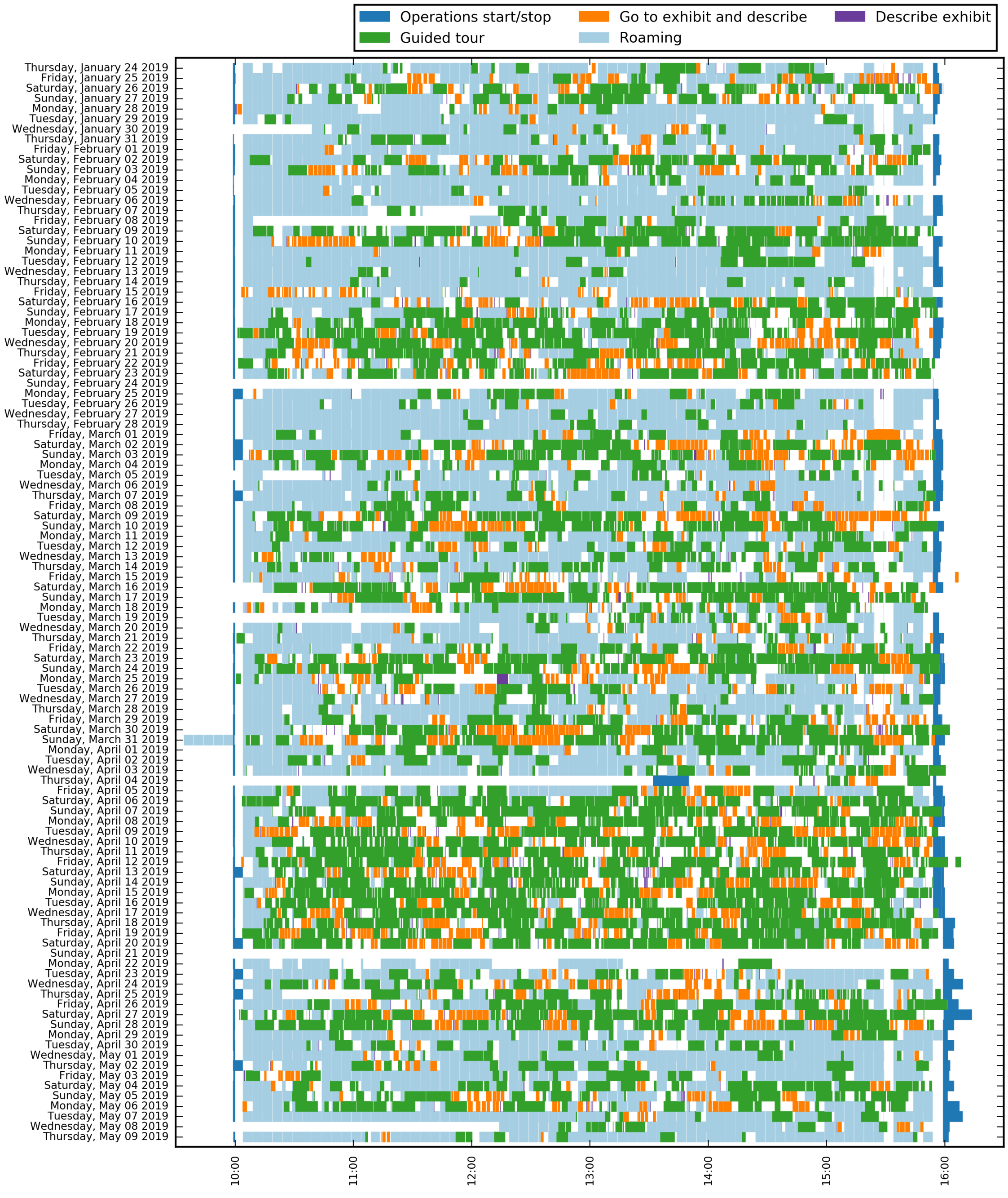


Figure 3.14: A plot of the tasks performed by the robot during deployment for nearly four months. White areas indicate that the robot is not performing any tasks. This can indicate that the robot is idle (e.g. when the robot is charging) or a failure has occurred.

Table 3.4: Navigation failures data.

Total navigation failures	1247
Tours with navigation failures	988
Average navigation failures per tour	0.16
Navigation failures helped	813

robot behaviours and capabilities have remained stable for most of the deployment, the variations in the number of user-demanded tasks over the different days must depend on several factors that are out of the robot developers' control, such as holidays, weather or the presence of other temporary exhibitions displayed at the museum. For example, a pattern can be observed from the figure where the tasks are more frequently demanded during weekends, as also discussed in the previous section. Moreover, the periods from the 16^{th} February to the 23^{th} February and from the 5^{th} April to the 20^{th} April, which saw a substantial increase in the number of user demanded tours with respect to the rest of the days, correspond to school term holidays periods in Lincolnshire (i.e. Spring Half Term Holidays: 16^{th} February - 25^{th} February, Spring Holidays: 5^{th} April - 23^{th} April). White spaces in the robot schedules indicate that the robot is charging (typically before 10:00 AM and after 4:00 PM), idle (after a user-demanded task) or that a failure has occurred.

Navigation failures

The navigation actions are amongst the most common causes of failures in our deployment in a public environment, as discussed in Subsection 3.5.1. In this context, a previous work by the author of this thesis [29] proposed a *Learning by Demonstration* framework that allows a robot to learn to detect navigation failures and generate recovery trajectories from demonstrations given by humans. This section, however, rather than describing the learning by-demonstration framework for navigation failures (which is not part of this thesis), focuses on analysing the navigation failures that occurred during the long-term deployment. In particular, in studying the navigation failures that have happened, it attempts to discern how many failures have caused a disruption of the interaction and how many failures have instead benefitted from the users' help to overcome the failure situation.

Navigation failures are caused by the navigation stack being unable to generate a path to reach the destination from the robot's current location. These usually happen when the robot is in a tight space, with a minimal distance from obstacles. Navigation failures are in fact, typically clustered in a few locations of the entire navigable space, where the shape of the environment makes failures more likely to happen.

Since the beginning of the deployment the robot has encountered 1247 failures during the 7876 guided tours it performed, hence there was a probability of 16 % of encountering a failure during a navigation action in the tour (see Table 3.4). The

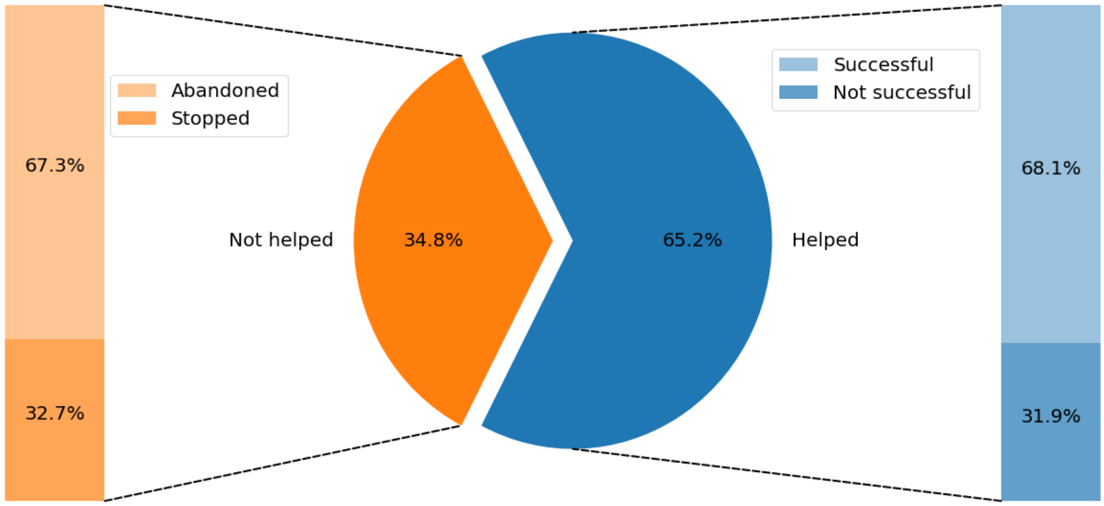


Figure 3.15: Pie chart showing the rate of navigation failures that were helped by the users when the robot asked them to. Out of all the helped ones, it shows the portion that turned out to lead to successful navigation; out of the not helped ones, it details the rate of people abandoning the robot or stopping the interaction.

number of tours with at least one failure is 988. In the event of a failure in navigation, the robot tries to return to a state where it can eventually resume the navigation by actively seeking the human help, as described in Subsection 3.5.1. Out of the 1247 navigation failures, in 813 failures(or 65.2% of the cases) the robot received help from the users. Figure 3.15 breaks down this statistic by showing how many of the helped failures were effectively recovered, with the robot successfully reaching the destination afterwards. Analysing the users' behaviour when they decided not to help the robot, the figure also shows that about 67% of the time users walk away from the robot requesting their help, or alternatively, they request to stop the guided tour. Finally, Figure 3.16 shows the trajectories generated by the users while dragging the robot to help it recover from a navigation failure. The plot confirms that failures are more frequent in tight spaces and shows high variability in how users have tried to help the robot.

Malfunctions and Other Disruptive Events

Lindsey, the robot, has been deployed in the museum for a total period of over 3.5 years to date. During this long time, however, several events and malfunctions have halted the robot's operations for various amounts of time. While some of such events have halted the robot operations for only a few days, others have caused significant disruption by stopping the deployment for months, as shown in Figure 3.17.

The following list outlines the most important of such occurrences, which are reported here in an attempt to inform future long-term deployments in public environments.

- **The robot falls into the floor.** During a repair of the museum building, the

Figure 3.16: Trajectories generated by the visitors helping the robot when unable to navigate. Note most locations where human help was given are clustered in the same few areas. Black areas are no-go zones for robot navigation.

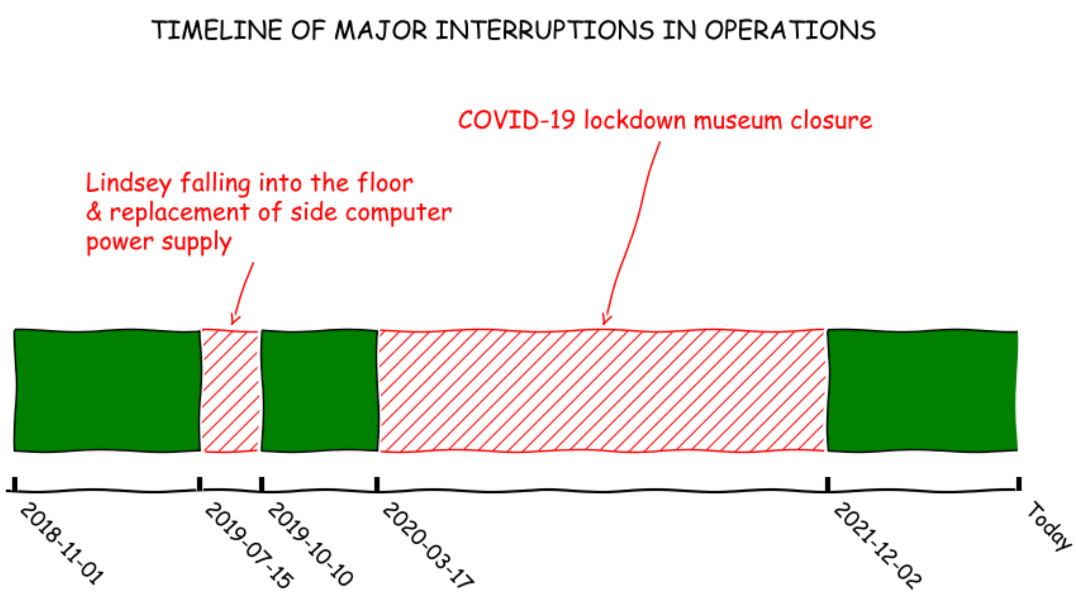


Figure 3.17: Timeline with Lindsey's deployment and the two major periods of interruption.

Figure 3.18: Lindsey fell into an opening of the floor of the museum.

workers opened the pavement floor, positioning a protective barrier to signal people about the work in progress. The robot, which was left free to roam around the museum, did not have the ability to detect barriers (like the one in Figure 3.18) or gaps in the floor. This incident halted the deployment for about three months. The robot detection abilities were enhanced by installing a downward-facing depth camera in the front side of the robot to detect obstacles and gaps in the floor during navigation.

- **Side computer power supply breakdown** The robot computing unit is composed of two computers, a central one which is directly powered by the manufacturer's power supply, and a side computer which hosts the robot's GPU (supported by an additional power supply unit). Given the high power requirement of the side computer, the power supply broke down, and a new and more powerful unit subsequently replaced it.

- **COVID museum closure.** This was the most disruptive event of the entire deployment, halting the robot deployment for more than one year and eight months.

- **Video card failure.** The video card of the main computational unit of the robot failed, leaving the robot without a touchscreen for interacting with users. The entire unit was replaced with a spare one that was readily available in the university lab. Therefore, this malfunction disrupted only one day of operations.

- **GPU fan failure.** The fan of the robot onboard GPU stopped working, and the GPU started overheating with time. This failure was detected thanks to the Sentor node (see Subsection 3.5.3), which alerted that the engagement prediction, running on the GPU, was not being published consistently over time.

- **Eyelid mechanism broken.** The robot has two mechanical eyes with a mechanism that allows it to gaze toward any location and close/open the eyelids. The mechanism of one of the eyes consistently wears off a gear that, over extended and continuous operations, eventually breaks.

3.7 Discussion

This section, drawing from the results of the analysis of the long-term deployment in Section 3.6, discusses the successes and limitations of the robotic system designed to deliver guided tours to museum visitors. Moreover, it sets out the basis for understanding how certain limitations are to be addressed in the future chapters of this thesis.

3.7.1 Autonomous Operations

From the description of the robot system design and the analysis of the performances in the museum, it should be clear by now that robot autonomy is a prerequisite for successful deployment in a public environment.

The data reported in Subsection 3.6.2 shows that the robot system was operative in the museum for 446 days, out of the total deployment time of 1209 days (corresponding to 3.3 years). In the remaining days, when the robot was not operative, there was either a robot malfunction –the robot was left stopped on the charging station while being repaired– or the museum was closed (for example, during the COVID lockdown). Despite these few occurennces along the way, detailed in Figure 3.6.2, the robot performed a total of more than 31 thousand tasks (including both the interactive and the routine tasks) and navigated autonomously more than a thousand kilometers to date. In line with previous work [40] on long-term deployments, the analysis reports the TSL and the A%. Both metrics highlight that the robot was available for autonomous operations for a good part of its time and was performing an interactive task with users for most of the deployment. Such results, considering the state of the art of other robot deployments, are an achievement in itself and leave us confident that Lindsey's robotic system is robust for the museum deployment as required by the **Objective 1** defined in Section 1.1.

3.7.2 Engagement with the Users

A second focus of the analysis in this chapter is studying the interactions between the robot and the museum visitors. The primary metric of interest when studying such a phenomenon is the user engagement with the robot and, more specifically, the robot-guided tours given that higher engagement in learning activities with a robot is linked to beneficial effects for the learners [37]. This metric is not directly accessible or measurable in the current setup and can only be estimated by proxy variables, such as the number of interactions and their duration as commonly done in previous work, such as [38]. User questionnaires –which can allow estimating the engagement by directly asking users to provide feedback– are not used in this work because the nature of the un-constrained long-term deployment did not allow having people on-site handing out forms to visitors at all times.

To fulfill **Objective 2**, the number of the interactive tasks started with the robot, their duration and their contextual information (such as the theme of tours) have been collected during the long-term deployment. The large number of interactions (i.e. more than seven thousand tours started in 446 days of operation) makes us confident that it was a popular attraction in the museum. From the analysis in Subsection 3.6.1 we also observe an increase in robot-guided tours during the weekends and school holidays, when more people usually visit the museum. The data also shows that people had a clear preference in the choice of the theme for the guided tour. However, this substantial amount of data from users' interactions are also evidence of the fact that the users' engagement is soon lost in interactions that are too extended in time. This is consistent with prior studies from similar deployments, where initial impressions of such robots are positive [48], but the interest fades throughout the interaction. This result could mean either that there is an intrinsic limit, of a few minutes, to the duration of human attention during guided tours or that the "robotised" tours are not engaging enough to keep the visitors' attention alive for longer. The former hypothesis implies that this threshold of the duration of attention would also be detected in human-guided tours. Human experience in such situations would suggest that this is not the case. While in a typical human-guided tour, visitors cannot typically stop the tour (but in severe situations) and they would not wander off due to compliance with social norms (avoiding impoliteness toward the guide, for example), in the robotic guided tours visitors are free to stop the tour whenever they like and do not appear to feel the urge of complying with human social norms toward the robot. The latter hypothesis, i.e. that the robot performs poorly in maintaining the visitors' engagement during the tours, implies that the robot's behaviours are not engaging enough. As with many robot deployments, there is a common (implicit) assumption that the robot is engaging by virtue of its presence alone. Our results are consistent with previous studies, such as [69], suggesting that this effect is not persistent over time, especially for more extended interactions.

From an alternative point of view, studies of human-guided tours in museums have provided recommendations for how human guides can give better (more engaging) tours [14], such as:

- tours should not resemble monolithic lectures, but they must be interactive;
- guides should facilitate audience contribution and engagement through questions and answers, also taking into account non-verbal features like eye movements and posture;
- guides should seek to secure the audience attention to inform and entertain them, encouraging them to orient to the feature under consideration;
- the audience should not be considerated as a whole, but the guide must take into account features of the single people, even personalising the experience;
- technologists need to create non-human guides with a similar level of sensitivity to the audience built-in.

Although primarily aimed at human guides, the above list gives valuable insights to roboticists for designing more engaging cultural experiences with robots. By testing these recommendations against the robot's tours, it becomes obvious how most (if not all) are not achieved by the current implementation. Therefore, it can be deduced that the static robot behaviours implemented are not able to effectively engage users throughout the interaction. This conclusion starts shedding some light on the **Hypothesis 1**, from Section 1.1, which will be eventually tested in the following chapters. However, in order to fully implement the recommendations mentioned above, there are several technological obstacles that must be overcome. These include the online detection of engagement, which is a substantial challenge, and the development of socially contingent robot behaviours to act on this information. This, in turn, suggests a range of technologies and techniques required to detect these features.

Motivated by this discussion, the Chapter 4 and Chapter 5 of this thesis will provide solutions to these challenges by presenting an approach for the online detection of user engagement from the robot point of view and to adapting robot behaviours to such engagement respectively. The approaches are supported by the long-term deployment in the museum, as the 'end users' would be directly involved in the learning process and the use PNP to specify robot behaviours that can be easily translated into stochastic policies, allowing learning.

3.8 Summary

This chapter presented the setup for the robot's long-term autonomous deployment by describing the museum environment, the robot hardware and the software system. The work is mainly focused on the aspects concerning the robot autonomy –i.e. the ability to remain available and operational for interacting with the users over time with minimal assistance required– and the way it interacts with the users.

A large amount of anonymised data was collected during the long-term deployment, lasting years to date, which allowed to analyse the performances for autonomy and user engagement. The analysis shows that the robot system was generally able to remain autonomous thanks to: the implementation of recovery strategies in case of failures, monitoring tools and the real-time notification of critical events. Failures, malfunctions and operationally disruptive events still happened due to the inherent complexity of the robot system and the fact that it has to deal with the events affecting the real world and the presence of humans in the unconstrained museum envirnemnt.

The robot was very popular in the museum, where it interacted with thousands of people and performed hundreds of guided tours daily. There are, however, some limitations to the ability of the robot to engage with its users, given that a large number of the interactions were interrupted (actively stopped or the users walked away) before their end. The findings from this analysis motivate us to explore methods to incorporate the users' states into the robot behavioural planning and to adapt the robot behaviour to the different user's preferences.

CHAPTER 4

The Engagement Model

This chapter presents a learned model that can assess the engagement level of the users interacting with a robot, as a scalar value, in real-time and from the robot's own camera. The regression model is trained end-to-end from a dataset collected at The Collection museum and annotated by independent coders. This chapter describes the data collection process, the data coding procedure, the engagement model design, and the evaluation of the proposed method in a real-world deployment.

4.1 Introduction

An important component of the framework designed in Chapter 2 for giving robots the autonomy to adapt their social behaviour online from the users interactions is a model of engagement that can provide a holistic assessment of the users group as a scalar value. Such evaluation are then used to establish a reward function for optimising the robot's behaviour through Reinforcement Learning (RL) methods.

Drawing from the experience of the initial deployment of Lindsey the robot in a public museum, with its analysis reported in Chapter 3, it can be established that one of the key challenges for robots interacting with humans is maintaining the users engagement alive during the entire duration of the interaction. Users typically exhibit high engagement levels at the beginning of the interactions by virtue of being in contact with a novel technology, but this fades over time when the robot's behaviours are static and do not take into account the users' state during the interaction [56, 47]. To address such shortcomings of current social robotics systems, the approach proposed in this thesis is to make the robot aware of the engagement state displayed by the users and to use such information to directly inform the robot behavioural planning. Therefore, a first step toward enabling robots to perform behaviours that are more engaging for users is to give them the ability to detect the users' engagement level during the interactions as a scalar value. With such an ability, a robot can then assess how its different behaviours affect the engagement level and, in turn, decide to execute actions expected to lead to higher engagements. The estimation of users' engagement is therefore considered an important step in the direction of automatic assessment of the robot's behaviours in terms of its social and communicative abilities, in order to facilitate *in-situ* adaptation and learning. In the context of RL, a scalar measure of engagement can directly be interpreted as a reinforcement signal that can eventually be used to govern the learning of suitable actions in the robot's operational situation and environment, as it will be demonstrated in Chapter 5. As a guiding principle (and indeed a working hypothesis), it is anticipated that a higher and sustained engagement with a robot can be interpreted as a positive reinforcement of the robot's action, allowing it to improve its behaviour in the long term.

The main challenge for assessing the engagement of humans during an interactive task with the robot, comes the fact that engagement is an inherently internal mental state which cannot be observed directly from the robot's perspective, or the human's perspective. Both humans and current robotic technologies have to resort to analyse the engagement from external cues from various sensor modalities (vision, speech, touch). However, in **Hypothesis 2** it is presumed that this assessment is intuitive from the human perspective thanks to their accumulated experiences in interacting with other people and thanks to the fact that they express engagement themselves, it remains challenging for robotic systems which typically lack the ability to learn from past interactions and to manifest engagement. Several research in different application domains of HRI [78, 10, 12] have focused on a measure of engagement to inform the assessment of the implementation for a specific use-case or to guide a robot's behaviour. However, the way engagement is measured and represented can vary in different contexts (as it will be discussed in Section 4.2) and it is probably not possible to find a generally applicable measure of engagement that readily lends itself to different social contexts.

Based on the observation that engagement as a concept is implicitly often quite intuitive for humans to assess, but inherently difficult to formalize into a simple and universal computational model, this chapter proposes to employ a data-driven machine learning approach, to exploit the implicit awareness of humans in assessing an interactive situation. Consequently, instead of aiming to comprehensively model and describe engagement as a multi-factored analysis, end-to-end machine learning is used to directly learn a regression model from video frames onto a scalar in the range of 0% to 100%. To train such a regression model, a rich annotated dataset is obtained from a long-term deployment of a robot tour guide in a museum.

Arguably, for a scalar engagement measure to be useful in actual HRI scenarios, a few requirements have to be fulfilled. In particular, the proposed solution should

- demonstrably generalize to new unseen people, environments, and situations;
- operate from a robot's point of view, forgoing any additional sensors in the environment;
- employ a sensing modality that is readily available on a variety of robot platforms;

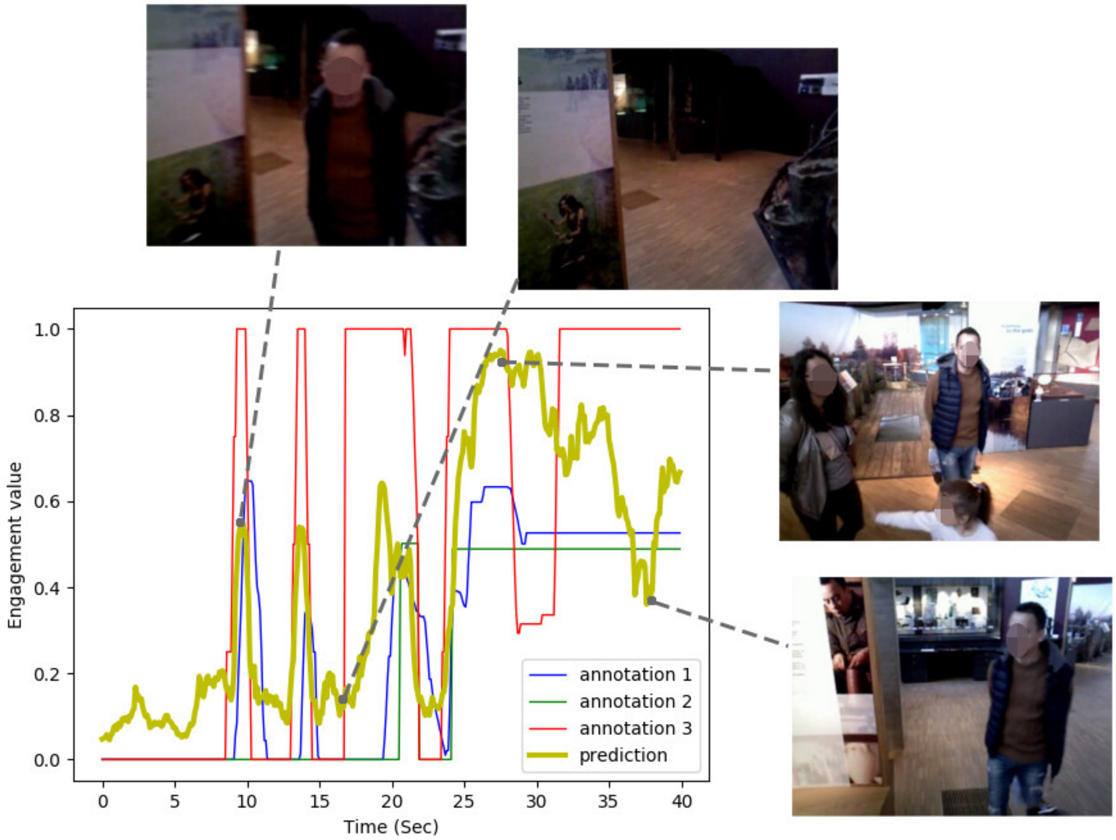


Figure 4.1: The continuous engagement annotations provided by the 3 coders with the engagement assessment prediction of the proposed model during a guided tour sequence. The images are recorded from our robot's head camera and are shown for only some moments of the entire interaction for illustrative purposes. Faces from the original dataset are blurred for anonymisation.

- have few additional software dependencies to maximize community uptake; and
- operate with modest computational resources at soft real-time.

Consequently, this chapter presents a novel engagement model, solely operating on a first-person (robot-centric) point of view and proves its applicability not only in the museum scenario but also on a publicly available dataset (UE-HRI) [12] without transfer learning or adaptation necessary. Results demonstrate that the model can operate at a 5 Hz frame rate on average GPU hardware typically found on robots. Figure 4.1 shows an example of the real-time engagement assessment during a real-world interaction with a group of people compared with the engagement assessments of from different human coders of the same interaction.

4.2 Related Work

Analysing the level of engagement of users during human-robot interactions is an important aspect in the field of social robotics. In the context of a learning environment, like the cultural context of a museum, it is known that a high student engagement generates better learning outcomes [72]. Moreover, engagement with a robot during learning activities has been shown to have a beneficial effect, facilitating communication and interaction [37]. These same results are also valid in HRI studies with people with disabilities [5]. While there is evidence that the presence of a robot, particularly when novel, is sufficient in itself for higher engagement in educational activities, e.g. [9], in interactions between individuals and the robot within a direct (social) interaction this effect typically fades over time, as it was observed in Chapter 3.

The facts that being able to elicit high engagement values during an interaction is beneficial for achieving high-quality outcomes and the knowledge that the sole presence of a robot is not sufficient for high engagements, motivates the study and development of technologies that can allow the assessment of users' engagement during the interactions with robots. Some previous works performed an assessment of engagement offline from video recordings of the interactions with the goal of evaluating the quality of the human-robot interaction and eventually of the robot's ability to engage [86, 10]. However, these works are not applicable to those contexts where the assessment should be performed online (i.e., while the interaction is taking place). In this work, engagement assessment should guide the robot's decision-making and therefore be performed online. For this reason, the approach followed is inspired by the works of Foster et al. [33], Castellano et al. [21] and Rich et al. [75] which all proposed models for detecting engagement in real-time from different sensors in the environment, or on the robot board, to inform the robot behaviours. These works however, all have their own different limitations which prevent their deployment on a mobile robot in an in-the-wild environment (such as a public museum). These limitations include being dependent on interacting with one single user, on the robot and the users being stationary in front of each other while interacting, or on the presence of additional sensors placed in the environment, other than those on the robot board. Differently, the requirements in this project are that the assessment should be performed from the robot's onboard sensors, with an arbitrary number of users and with both the robot and the users moving freely in the environment.

In the search for a suitable methodology for assessing engagement and to guide the design of novel approaches, it is important to first study how engagement with a robot, or a technology, is defined in the field and what different aspects of it should be considered. However, despite a large number of studies on user engagement in HRI there is yet to be found a universally agreed definition [36] for it, and consequently, many different approaches were proposed in previous work for evaluating engagement in interactions.

In the following subsections, first the definitions of engagement available in the literature will be analysed, and then previous approaches at characterising user engagement in HRI will be reviewed and discussed in the context of the requirements of the present scenario.

4.2.1 Definitions of engagement

Engagement is a term used to indicate a phenomenon that is naturally easy to understand and recognise by humans. However, in trying to give it a conceptual definition, one finds that as a concept, it is more convoluted than what it intuitively seems. The literature when attempting to define *what* engagement is, has taken two different approaches: the first, viewing engagement as a process taking place during the interactions, and the second, defining it as a metric of the interaction quality, which value can be estimated from observations.

Within the first group of works, Sidner et al. [86, 87] describe engagement as 'the process by which individuals in an interaction start, maintain and end their perceived connection to one another' and 'it combines verbal communication and non-verbal behaviours, all of which support the perception of connectedness between interactors'. O'Brien et al. [70] define engagement with technology as 'a process comprised of four distinct stages: point of engagement, period of sustained engagement, disengagement, and re-engagement'; with the process being characterised 'by the presence of multiple attributes that vary in intensity depending on a combination of user and system attributes that emerge during the interaction'. The attributes considered are 'challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect'.

Between the works that have taken the assumption that engagement is a metric of the interaction, Peters et al. [71] defines engagement as 'the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction'. They identify two different moments of the interaction which are most relevant for assessing engagement: at the moment of starting a communicative interaction (to assess the possibility of engagement in interaction) and when the interaction is ongoing (to check if engagement is lasting and sustaining conversation). According to Salam et al. [81], engagement is a social dimension that can be seen as 'the measure of the intention-to and the quality-of interaction as perceived by the user'.

For the purpose of the work presented in this thesis, the latter idea is pursued, i.e. that engagement is a measure of the interaction quality and can be evaluated during social interactions, rather than attempting to detect the different engagement stages that are part of the engagement process. This choice reflects our goal of viewing user engagement as a feedback to the robot behaviour, i.e. as a metric of the quality of the interaction which is (partly) affected by the robot's actions.

4.2.2 Characterisation of engagement

Despite the lack of a common definition of engagement, an exceptionally large number of works have proposed methods for its characterisation in Human Robot Interaction (HRI).

Peters et al. [71] identify engagement and interest as causal factors of attention and devise an algorithm based on gaze for detecting engagement in interactions. In addition to interest and attention, Castellano et al. showed that an affective component (e.g. valence) can be integrated into the characterisation of engagement from the perception of the user's facial features [20] and from the robot's own affective expressions [21]. In both human-human and human-robot interactions, gaze has been identified as particularly significant for determining engagement levels in an interaction, for example looking at the work of Rich et al [75] and Holroyd [43]. Therefore, gaze forms an essential behavioural cue when assessing engagement, e.g. [86, 10]. Lemaignan et al. [58] instead of trying to directly define and detect engagement, recognising that it is a complex and broad concept, introduces the concept of "with-me-ness", which is the extent to which the human is "with" the robot during the interactions and is solely based on the human gaze behaviour.

Beyond only non-verbal behaviours, Foster et al. [34] attempt to estimate the customers' engagement state from the audio-visual sensors data of a robot bartender. Sidner et al. [86, 87] also combine verbal communication (user utterances and sound location) with non-verbal behaviours to 'support the perception of connectedness between interactors'.

Moreover, context has also been identified as being of importance, as it manifests in the specifics of the task and environment, as well as the social context [19]. For example, Michalowski et al. [66] propose a simple model to infer engagement for a robot receptionist based on the person's spatial position within some predefined areas around the robot. Salam et al. [82] attempt to predict the engagement of one entity in a multiparty interaction relying only on the features of rest of the group, showing that engagement, and the features needed to detect it, changes with the context of the interaction [81]. Similar results from Leite et al. [57] show that the prediction of disengagement in one-person interactions versus group interactions relies on different set of features that can be observed. However, they report that a model trained only on group interaction data might perform reasonably well also in scenarios with a single user.

These examples suggest that multiple, overlapping, and likely interacting observable features are involved in the characterisation of engagement, including at different timescales, from the longer-term context to short interaction-orientated behaviours that nevertheless impact social dynamics and which humans are particularly receptive to [31]. Based on these observations, in this work engagement is explicitly not modeled on specific observational cues, but a general model is proposed that can potentially integrate multiple features (such as gaze, context and group size) at different timescales.

In addition to explicitly cue-centred approaches, more recently, attempts have been made to leverage the power of machine learning to discover the important overtly visible features with minimal (or at least sparse) explicit guidance from humans (through cue identification, for example). For example, Foster et al. [33] compare a hand-coded rule-based baseline to learned classifiers for detecting when clients of a robot bartender were seeking engagement with the robot. They found that in offline evaluation the learned classifiers performed significantly better, but in an online evaluation no significant difference between the conditions was found. Won Park et al. [98] use an active learning approach with Deep RL to automatically (and interactively) learn the engagement level of children interacting with a robot from raw video sequences. The learning is incremental and allows for the real-time update of the estimates so that the results can be adapted to different users or situations. The DQN is initially trained with videos labelled with engagement values. In other work, Youssef et al. [13] trains DL models to classify *Signs of Engagement Decrease* events from low-level features such as distance from the robot, gaze, head angles, face Action Units [6] and speech features, extracted by the robot's modules. Rudovic et al. [79] investigates the performance of a DL model, called CultureNet, to specifically estimate the engagement of children with autism coming from different cultural backgrounds by studying the performance across the multicultural data, although this is based on a dataset of images of children's faces rather than real-time data.

These deep learning methods have the advantage that the constituent features of interest do not have to be explicitly defined *a priori* by the system designer, instead, only the (hidden) phenomenon needs to be annotated; engagement in this case. Since social engagement within interactions is readily recognised by humans based on visible information (see discussion above), human coding of engagement provides a promising source of ground-truth information. Indeed, in this context, Tanaka et al. [91] employed human coders to assess the 'quality' of observed interactions, demonstrating good agreement between coders on what was a subjective metric.

Taken together, the literature indicates that while a precise operational definition of engagement may not be universally agreed, it seems that more holistic perspectives may be more insightful. It is likely that, while gaze is an important cue in making this assessment, other contextual factors influence the interpretation of engagement. Given that humans are naturally able to assess engagement in interactions accurately, it seems that one promising possibility would be to leverage this to inform automated systems directly. Based on the approaches in the literature reviewed, the engagement model proposed in this thesis is trained to recognise engagement from raw image sequences using a Deep Learning (DL) model, to potentially integrate any visible features related to engagement, and from human annotations of engagement. This enables to harness the coders intutive understanding of user engagement, without the need to rely on specific engagement definitions or features.

4.3 Preliminaries

In this section, the methodologies that are at the foundations of the proposed engagement assessment model are briefly presented. Since previous work has shown that deep convolutional networks can be used effectively to predict social signals from human-robot interaction videos, e.g. [79, 74], an architecture composed of a convolutional layer, to extract features from the raw data, and a recurrent layer, to integrate these features among the temporal dimention, is desgined.

4.3.1 Artificial Neural Networks

An Artificial Neural Network (ANN), also referred to as simply Neural Network, is a computational method inspired by the functioning of the networks in biological brains. The networks are formed by a collection of nodes and directed edges connecting the nodes, forming a directed graph structure. Each node takes in input the values arriving from the incoming edges and performs a simple operation (typically non-linear) to aggregate such values into a single scalar that is sent out as output. Therefore, nodes can communicate through the edges connecting them to other nodes, i.e. through the edges that send the node's output to other nodes (or to themselves) as an input. Edges have a weight each which can excite or dampen the value that is transmitted in it. In most cases, the nodes in the network are structured in layers where the information flows always in the same direction, from nodes in the *(i)*-th layer to the *(i+1)*-th layer (see Figure 4.2). In such structure usually one can identify an *input* layer (the *0*-th layer) –which nodes contains the values of external quantities observed–, one or more *hidden* layers –processing the information coming from the input– and an output layer –which makes available the results of the network's computation. When the network has multiple layers it can be called a Deep Network –consequently the word Deep Learning (DL) refers to the use of Deep Networks for learning–, opposed to a *Perceptron* that is formed by a single layer.

Learning in ANN happens by means of adjusting the weights of the edges so that a *cost function* is minimised. The cost function is usually defined over the output vector. With the *supervised learning* paradigm for example, there is a dataset containing the input and output pairs that the network need to match and the cost function is defined as the error between the given output (i.e. the ground truth) and the output of the network. When using other paradigms, like *unsupervised learning* and *self-supervised learning*, the output value is not provided by the dataset. In such paradigms the cost function is either defined based on the task and prior assumptions about the model (e.g. maximising the mutual information between inputs and outputs), or it can be defined as minimising the difference between the

[1] Contributors to Wikimedia projects. (2022, June 10). Artificial neural network - Wikipedia. Retrieved from `https://en.wikipedia.org/w/index.php?title=Artificial_neural_network&oldid=1092513725`

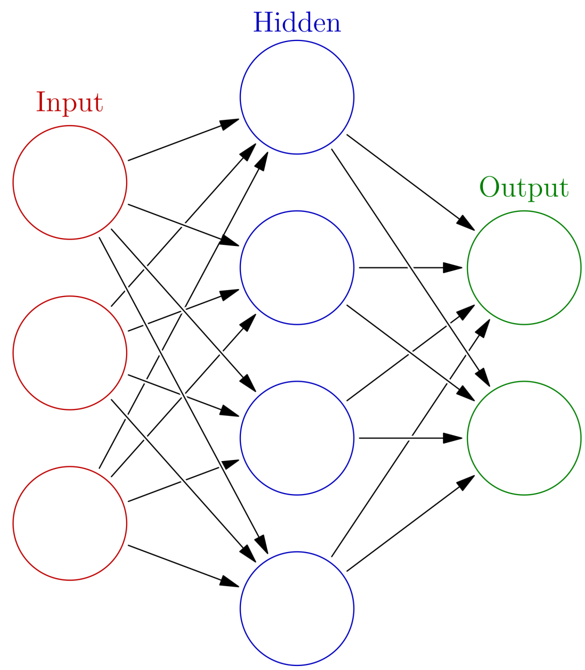


Figure 4.2: An Artificial Neural Network with nodes and edges organised in *input*, *hidden*, and *output* layers.[1]

network output and an output obtained from the data itself (e.g. minimizing the output difference between consecutive image frames).

Convolutional Networks

Convolutional Neural Network (CNN) is a type of ANN which became a foundation model for computer vision applications after it was shown by LeCun et al. [54] that, in combination with back-propagation [80] for learning the network weights, it performs significantly better than other approaches at hand-written recognition tasks.

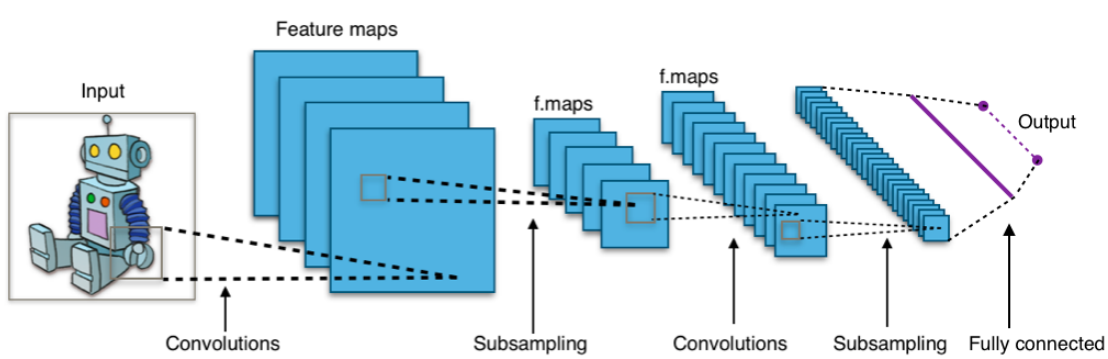


Figure 4.3: The structure of a typical Convolutional Neural Network for vision applications.[2]

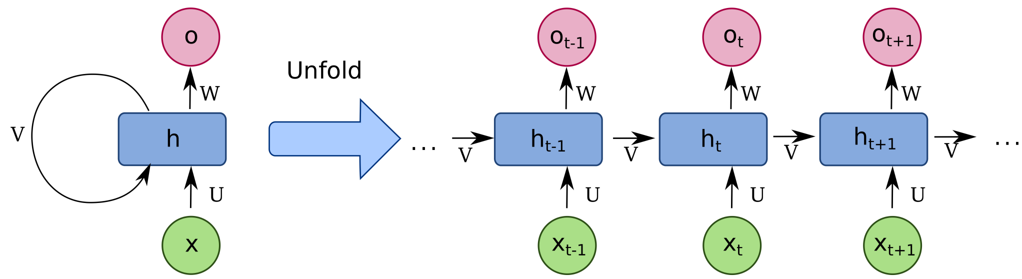


Figure 4.4: (left)A Recurrent Neural Network without the temporal dimension; and (right) the same network unfolded over multiple time steps.[3]

Differently than a traditional ANN, the layers of convolutional networks are not simple collections of nodes that feed from the output of all the previous layer's nodes, but rather each node only collects the input from a local region of previous nodes (see Figure 4.3). Even if a node only looks at a local region, it's operation is performed on the entire previous layer by sliding over all the possible local regions. This operation creates the so called feature maps, i.e. collections of nodes representing a specific feature of the previous layer. The layers performing such local operations are called *convolutional layers*. Another important feature of CNN are *pooling layers*, which perform a subsampling of the previous layer (e.g. by performing max pooling) effectively reducing the size of the feature maps. Typical convolutional networks are formed by a sequence of convolutional and pooling layers, followed by one or more fully connected layers.

Recurrent Networks

Recurrent neural networks are a type of *ANN* where some of the edges between nodes are organised to represent the temporal dimension, i.e. an edge that connects a node at time *t* with the same node at time *t+1*. These networks are called "recurrent" because when removing the temporal dimension the recurrent edges become feedback connections, which start and end on the same node (see Figure 4.4).

One of the most common type of recurrent network is the Long-Short Term Memory (LSTM) which was invented by Hochreiter and Schmidhuber [41] in 1997. In a LSTM the nodes are organised in units containing a *cell* (i.e. the internal memory), an *input gate* (which controls the amount of input allowed in), an *output*

[2]Image from contributors to Wikimedia projects. (2022, June 22). Convolutional neural network - Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Convolutional_neural_network&oldid=1094374344

[3]Image from contributors to Wikimedia projects. (2022, June 16). Recurrent neural network - Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Recurrent_neural_network&oldid=1093481842

gate (controlling the amount allowed out) and a *forget gate* (allowing to forget the current cell value).

4.4 The TOur GUide RObot (TOGURO) Dataset

For almost the entire duration of the deployment the data about the robot operations, its internal state and the sensor data, among others, have been collected daily into a database. In this chapter, for the purpose of assessing users' engagement only a subset of the entire database collected is taken in consideration. Such subset consists of the video recordings from the RGB-D cameras of the robot collected during the interactions. The work and data recording exercise have been approved by the University of Lincoln's Ethics Board, under approval ID "COSREC509". The ethical approval does not allow the public release of any data that can feature identifiable persons, in particular video data. For this reason, in the rest of this chapter pictures from the dataset have been anonymised. The data utilized in the analysis reported in this section spans the date range between the 24th January 2019 (the day on which data recording started) and the 17th March 2020, with data collection remaining ongoing.

4.4.1 Dataset Collection

The TOur GUide RObot (TOGURO) dataset was collected from the two cameras mounted on the robot's body and head, each providing a stream of RGB and depth frames. Each video stream was recorded from the moment the user started a *guided tour* or a *go to exhibit and describe* task and continued until its termination, either when the end of the interactive task was reached or because of users stopping/abandoning it. The recordings therefore do not include the initiation of engagement phase. *Describe exhibit* tasks were excluded from the data collection because of their very short duration. In order to minimise the storage required for such amount of videos, the different frames coming from the various cameras are collected, compressed and stored as MPEG videos on-the-fly while the interaction is taking place. To be able to reconstruct the frame-by-frame alignment between the different video streams, the ROS timestamp of each received camera frame message is stored alongside the frames themselves.

The participants were aware that the robot was recording data during the interactions (by means of visible texts on the robot's display and leaflets), however they were not informed of the purpose of the recording, i.e. the assessment and analysis of users engagement, in order to not bias their behavior. In total, the robot collected 3106 distinct videos for a total duration of about 10 days and 16 hours of recorded interactions for each camera stream. Note, however, that only a subset of this total data was coded and used for training and evaluation of the proposed model, as described in the following section.

Given that the museum in which the robot is deployed is a public space openly accessible to anyone, the interactions between the robot and the museum's visitors are completely unstructured. People walking in the gallery are allowed to roam around the collection or to interact with the robot. When they choose to do so, they do not receive any instruction about how to interact with it explicitly and are not observed by experimenters.

4.4.2 Dataset Coding

In order to provide a ground truth for the assessment of robot-centric group engagement, a subset of the TOGURO dataset was manually coded. As noted previously, given that there is not a universally accepted operationalized definition of engagement, a human observer response method is employed in the present work, following the prior application of a continuous audience response method [91].

Three annotators took part in the coding process: each was familiar with the robot being used and the interaction context. The annotators were students at the University of Lincoln, who knew each other before the study and were not remunerated for the activity. They were instructed to provide engagement scores as scalar values in the range $[0, 1]$ to reflect the following measure: *"Situate yourself as the guide (i.e. the robot) carrying out the interaction and looking through the cameras. How much (do you feel) the people are being engaged by the interaction with you?"*. A set of exemplary cases was also provided:

- Whenever the person is looking at the screen or at the robot head: engagement is HIGH;
- When the person is looking at the exhibit (the exhibit is typically behind the robot): engagement is HIGH;
- When the person is attending the tour but annoyed, continuously looking around, or looking at the phone: the engagement is MEDIUM;
- When the person is not attending the tour (e.g. far from the robot, oriented with the back toward the robot, talking to other people): engagement is LOW;
- When the person is not in the camera field of view: engagement is LOW;
- When the face of the people are not completely visible, do not immediately classify engagement as if the people were outside the FOV but try to guess their engagement value;

where HIGH, MEDIUM and LOW do not identify a precise discrete value but they are an indicator of the scalar value range. Reflecting on the nature of the interactions in the museum, the score provided by the coders takes into account the situations where a user diverts its attention from the robot but remains essentially engaged in the task by looking at the exhibits.

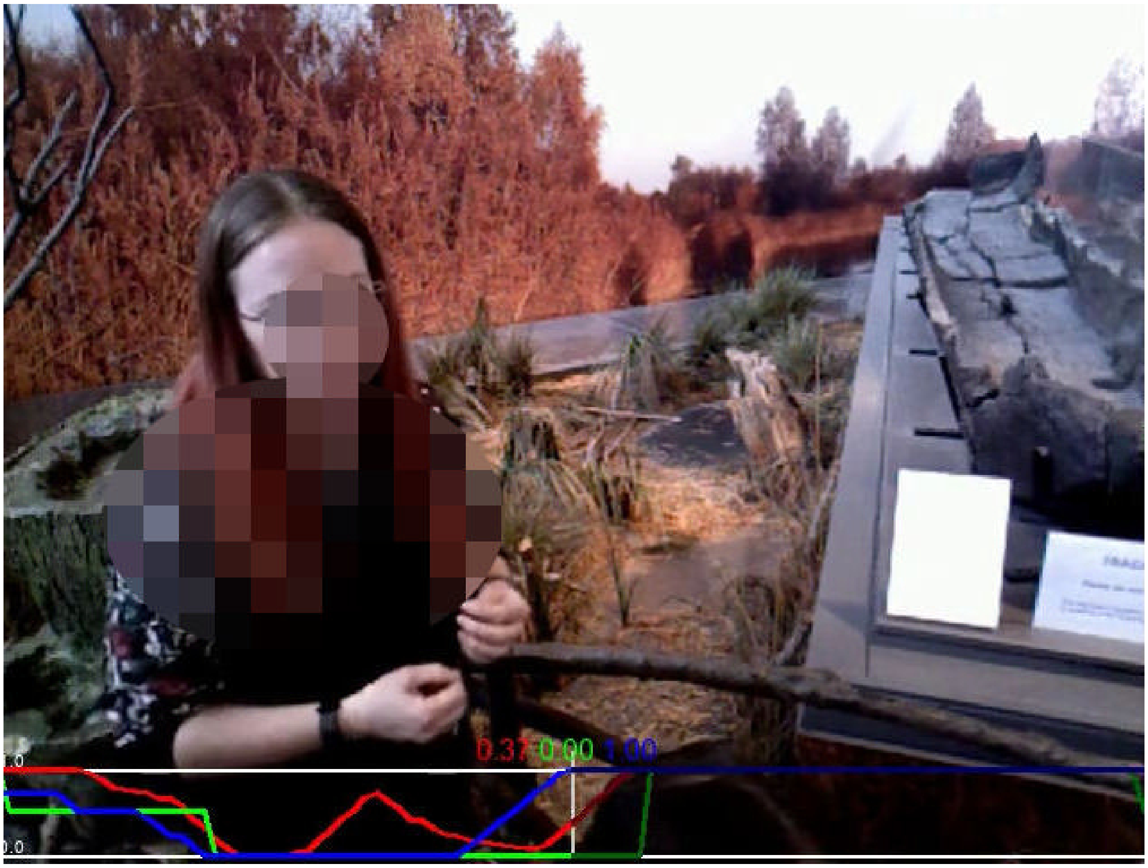


Figure 4.5: One frame from a video in the TOGURO dataset recorded from the robot's head camera during a guided tour. The red, green and blue plots at the bottom of the frame represent each a distinct annotation sequence. Face from original dataset blurred for anonymisation.

The annotations were performed over only the RGB stream of the robot's head camera, rather than considering all the four video streams available from the collected data. Similarly to [91], the annotators were asked to indicate in real-time how engaged people interacting with a robot appeared to be in the video captured by the robot (e.g. Figure 4.5). They operated a dial using a game-pad joystick while watching the interaction videos using the NOVA annotation tool[4] [8]. This procedure allowed the generation of per-frame annotations of the provided videos, with very little time spent on software training (around 20 minutes per annotator) and on the annotation process itself (not more than the duration of the videos).

Three subsets of the overall dataset were randomly drawn and assigned to the annotators. The subsets were partially overlapping in order to enable an analysis of the inter-rater agreement for assessing the reliability of the essentially subjective metric, but also to maximize annotation coverage of the dataset. As indicated in Table 4.1, the total length of the annotated data was over nine hours, with 3 hours 27m of overlap between the annotators (resulting in 5 hours 50m of unique videos

[4] https://github.com/hcmlab/nova

Table 4.1: Video annotations by annotator (coder): unique indicates the length of video coded by a single coder

Coder	# Videos	Tot Duration
`Coder1`	66	3h 59m
`Coder2`	40	2h 55m
`Coder3`	40	2h 23m
Unique	94	5h 50m
Total	146	9h 17m

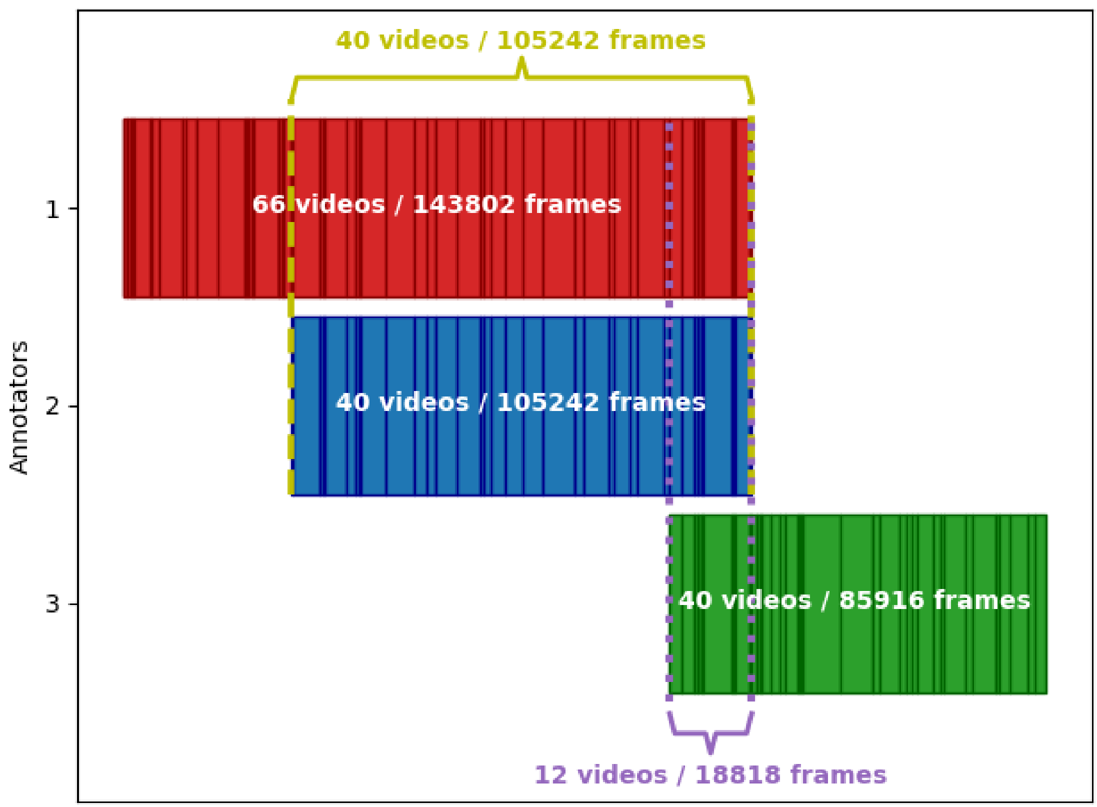


Figure 4.6: The videos annotated by the independent coders showing the amounts that have been annotated by 1, 2 or all 3 coders. The vertical darker lines in the bars separates the different videos. Note that the videos are stacked one next to the other for purely visual purposes; they have been randomly sampled from a larger pool and therefore are not temporally correlated.

Table 4.2: Spearman's Correlation ρ at different smoothing constant values S. The significance p-value < 0.001 and sample size $n \geq 89$ for all coder pairs and smoothing constants.

Coders Pair	S **[sec]**	ρ
`Coder1` $\leftrightarrow$ `Coder2`	1	0.71
	5	0.77
	10	0.79
	26	0.78
`Coder1` $\leftrightarrow$ `Coder3`	1	0.49
	5	0.5
	10	0.52
	26	0.65
`Coder2` $\leftrightarrow$ `Coder3`	1	0.48
	5	0.5
	10	0.53
	26	0.72
Average	1	0.56
	5	0.59
	10	0.62
	26	0.72

annotated). The amount of annotated data, with the fractions that multiple coders annotated, is also depicted in Figure 4.6. The three annotators coded 96 unique videos with a total of 146 videos (including repeated annotations) for a total duration of 9 hours and 17 minutes. In total, the annotated video set featured 227 people, of which 53.74% (122) were females and 46.26% (105) males, 60.79% (138) where adults and 39.21% (89) minors. The composition of each group of people interacting with the robot is very diverse; on average each videos features 2.41 people ($min = 0, max = 9, \sigma = 1.56$), 1.32 females ($min = 0, max = 6, \sigma = 0.89$), 1.14 males ($min = 0, max = 5, \sigma = 1.26$), 1.5 adults ($min = 0, max = 5, \sigma = 0.97$) and 0.96 minors ($min = 0, max = 6, \sigma = 1.14$).

Coding evaluation

The annotated engagement rating is a continuous scalar value given for every frame of the video data. As such, Spearman's rank correlation [88], indicated with ρ, is used to assess inter-rater agreement. Previous work has usually employed the Cohen's Kappa statistic [62] for determining the reliability of multiple coders for categorical annotations; however, this is not applicable in the current context where the annotations are continuous scalar values. Table 4.2 shows the correlation value for each pair of annotators. Since a scalar annotation value is provided for every frame, at a framerate of 10 frames-per-second, directly comparing the per-frame values does not provide a meaningful evaluation. Therefore, the correlation of the continuous values is evaluated multiple times for different smoothing windows with constant window values in the range $[0.1s, 40s]$ (see Figure 4.7(A)). Table 4.2 provides a summary of these, with overall mean agreement rates at selected representative values of the smoothing constant. While there is variability in the between-coder agreement, mean values of ρ vary in strength from moderate to strong (0.56 to 0.72). In this regard, there is a trade-off to be made between the smoothing constant size and the apparent agreement between the coders: a larger window size reduces the real-time relevance of the engagement assessment, even though the agreement over the extended periods of time is greater than in the comparatively shorter windows. Overall, these results indicate that the use of the independently coded data can be considered reliable in terms of the highly variable and subjective metric of engagement.

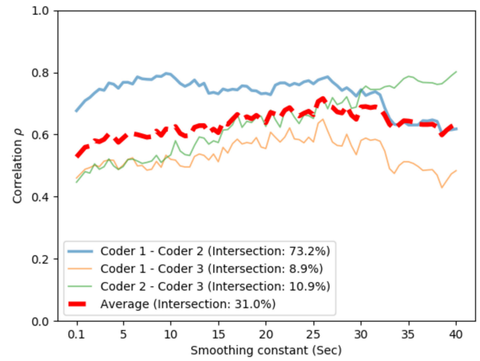
To further analyse the agreement between the coders, the Spearman's rank correlation is computed again separately for different interaction conditions which separates the dataset into the following four subsets:

- *single-party* interactions, i.e. with less than 2 people (22 unique videos);
- *multi-party* interactions, i.e. with at least 2 people (72 unique videos);
- *adults-only* interactions, i.e. where all the users are adults (38 unique videos);
- *adults-and-minors* interactions, i.e. where there is at least 1 minor (54 unique videos).

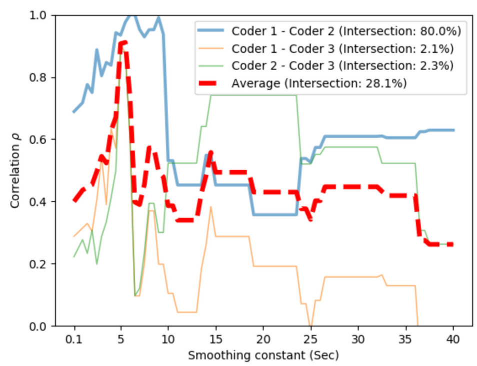
Figure 4.7(B-E) shows the correlation value at various timescales for each of these conditions. It can be observed that for the *single-party* and *adults-only* conditions the difference in correlation between coders is more sensitive to variations of the smoothing window size than in the other conditions. The correlation is in fact very high for some specific values of the constant, and average to low otherwise. In these two conditions the variation between different pairs of coders is also more pronounced. Such observations suggest that the assessment of engagement for single parties and in groups with only adults is generally more challenging than for the interactions with multiple people (whether those are only adults, or adults and minors). It is important to notice that the coders were not instructed to use different coding strategies for each interaction condition but only to code the overall group engagement as explained above.

4.5 The engagement regression model

Having established a ground-truth for the level of users' engagement, provided by the human-coded TOGURO dataset, a Deep Learning (DL) approach is proposed for the estimation of such engagement in real-time from the image sequences coming from the robot camera. The model is trained end-to-end based on the annotated videos

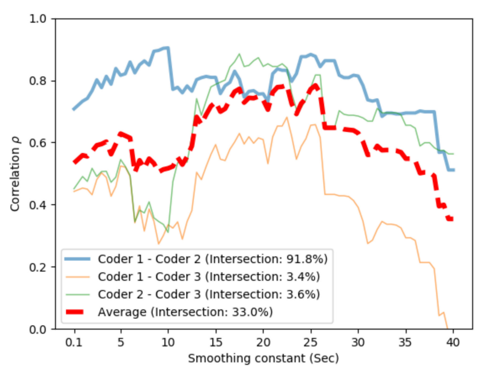


(A) Entire dataset.



(B) Single-party interactions.

(C) Multi-party interactions.



(D) Adults-only interactions.

(E) Adults-and-minors interactions.

Figure 4.7: Spearman correlation averaged over coder pairs and weighted by the overlap rate. Values are reported over different smoothing constants S and for different interaction conditions.

of the TOGURO dataset featuring people interacting with the robot. It should be noted that this approach does not attempt to model the individual humans that are in the robot's point of view but it provides an overall holistic engagement score, following the same principles applied during the coding procedure described above.

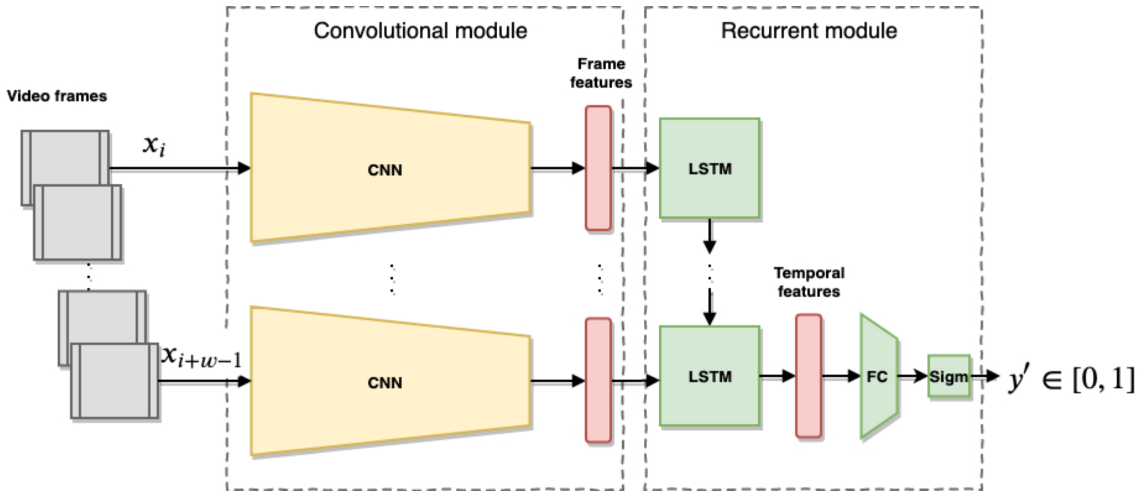


Figure 4.8: Overview of the proposed model. The input is a video stream of interactions between the robot and humans collected in w size intervals. The frames x_i are passed through the pre-trained CNN (ResNet), producing a per-frame feature vector which is then passed sequentially to the LSTM network. After w steps, the LSTM produces a temporal feature vector which is passed to a FC layer with sigmoid activation to produce an engagement value y for the temporal window.

The network architecture, depicted in Figure 4.8, is composed of two main parts: a convolutional module that extracts frame-wise image features, and a recurrent module that aggregates the single frame features over a time window to produce a temporal representation of the scene. The convolutional module used is a ResNetXt-50 Convolutional Neural Network (CNN) [99] pre-trained on the ImageNet dataset [50]. The use of a pretrained ResNet as a general purpose feature extractor model, in combination with a final task-dependent module trained ad-hoc, is an established practice such as in [79]. The single frame features, with dimension 2048, are collected after the activation function of the last fully connected layer, just before the final softmax layer of the network. The recurrent module is a single layer Long-Short Term Memory (LSTM) [42] with 2048 units followed by a Fully Connected (FC) layer of size 2048×1. The LSTM takes in input a sequence of w frame features coming from the convolutional module and produces, in turn, a feature vector that represents the entire frame sequence to capture the temporal behavior of humans within the time window w. The temporal features are passed through the FC layer with a sigmoid activation function at the end to produce values $y' \in [0, 1]$ The recurrent module is trained from scratch in the experiments to predict engagement values from the provided annotation values, while the pre-trained CNN module is fixed.

The proposed framework is implemented in Python using the Keras library[5] and is freely released as a ready-to-use ROS package to the HRI community [6].

4.6 Experiments

The model presented is trained and tested on the TOGURO dataset and, for the purpose of assessing generalisation in different settings, on the UE-HRI dataset.

4.6.1 TOGURO Dataset Processing

The annotated set from the TOGURO dataset that has been processed for the experimental evaluation is composed of 94 videos for a total duration of 5 hours and 50 minutes of interactions, as discussed in Subsection 4.4.2. For each video, the annotation from a randomly chosen coder is used when multiple annotations are available (see Figure 4.6), to avoid repetition in the data and to do not bias the model toward those videos that have been annotated multiple times. Each video is then randomly assigned to either the training, test or validation set with a corresponding sampling probability of 50%, 30%, and 20%, respectively, so that different frames from the same videos do not appear on multiple sets. This mechanism avoids testing and validating samples that are closely correlated with data seen during training, hence reporting a biased evaluation. Each video V_k, composed of I_{V_k} frames $x_i \in V_k$ for $i \in 0, \ldots, I_{V_k}$, has an associated array of annotations $A_k = [y_0, \ldots, y_{I_{V_k}}]$, also of dimension I_{V_k}. From all the videos in each set (training/test/validation) all the possible sequences of w consecutive frames $X_i = [x_i, \ldots, x_{i+w-1}]$ are extracted to form the input samples for the model. Therefore, each sample X_i overlaps by $w - 1$ frames with the sample X_{i+1} from the same video. To each X_i there is a corresponding label $y_{i+w-1} \in A_k$, that associates the sequence of frames to the engagement value perceived at the end of the sequence. In this way, each engagement value is associated with the temporal window of frames leading to it.

After the pre-processing phase over the annotated dataset, a total of 93 271 training samples, 72 146 test samples and 44 581 validation samples were obtained. Each frame is reshaped to size 224×224, and normalised before being fed to the network.

4.6.2 Training and Evaluation

For training and evaluation, the window size w is set to 10 frames in order to have a model that gives evaluations of engagement in a relatively short time (i.e. after 1 second). Even though more temporally extended time windows would provide more coherent ground truth values among the different annotators, as discussed in Section

[5]Team, K. (2022, March 01). Keras: the Python deep learning API. Retrieved from https://keras.io

[6]https://github.com/LCAS/engagement_detector

4.4.2, it was chosen to sacrifice some accuracy in favour of increased real timeless of the model predictions.

During training, the weights of the CNN module, which is already pre-trained, are kept frozen while the LSTM module is fully trained from scratch. The model is trained to optimise the Mean Squared Error (MSE) regression loss between the prediction values y'_i and the corresponding ground truth values y_i using the Adagrad optimisation algorithm [30] with an initial learning rate $lr = 1e-4$. At each training epoch, 20% of the training set is uniformly sampled to be used for training with a batch size $bs = 16$. The uniform data sampling of the training data is performed in order to reduce training time and limit overfitting [32]. The model was trained for 22 epochs using early stopping after no improvement in validation loss.

The evaluation of the model is performed on the entire test set, which is composed of samples from videos never seen in the training phase. Additionally, the performances are evaluated on the *single-party*, *multi-party*, *adults-only* and *adults-and-minors* portions of the test set separately in order to understand how the differences in groups affects the engagement estimation of the model. In these last evaluations, the same model learned from the entire training set is used rather than models trained from different portions of the data for each condition.

4.6.3 Evaluating Generalization

In order to assess the generalisation capabilities of the trained model over different scenarios featuring people interacting with robots, the performances are evaluated as the model's ability to detect the start and end of interactions over the UE-HRI dataset [12].

The UE-HRI dataset

Similarly to the TOGURO dataset, the UE-HRI dataset was collected during the public deployment of a social robot (i.e. SoftBank Robotics' Pepper) featuring interactions in-the-wild with groups of users. Moreover, the human-robot interactions are mediated through a touchscreen interface and speech. However, while in the scenario presented in this thesis the robot moves around the museum together with the users as it explains the various exhibits, the Pepper robot stands in a fixed position in the room interacting with people entering its own engagement zone by asking them questions and showing applications from the touchscreen. Moreover, the camera's position on the Pepper robot is different from that of the head camera on Lindsey (from which the TOGURO dataset was collected). Lindsey's camera is also in a fixed position during the interactions, while Pepper's camera moves with the robot's head movements. Despite such camera movements, the robot directs it's head towards the people, often shifting from one person to another. Therefore, their position in the camera view allows the estimation of the user's engagement from the robot's point of view.

Table 4.3: Model performance on our TOGURO Dataset. ρ measures the correlation between the predictions and the ground truth values with smoothing factor $S = 1$ [sec]. Prediction time is relative to the GPU GeForce GTX 1060 used to carry out the assessment.

Test loss (MSE)	**Correlation ρ**	**Prediction time**	**Memory usage**
0.126	0.634	$t <= 0.2$ sec	5.4GB

The videos are accompanied by annotations of start/end of interactions and various signs of engagement decrease (Sign of Engagement Decrease (SED), Early sign of future engagement BreakDown (EBD), engagement BreakDown (BD) and Temporary Disengagement (TD)). These annotated signals are associated with cues of verbal/non-verbal behaviours of the users and other various features, like the users' position. In total, the dataset features 54 interactions with 36 males and 18 females, where 32 are mono-users, and 22 are multiparty.

Evaluation Procedure

For a fair comparison with the proposed method, here it is evaluated the ability of the model to distinguish between the moments during which an interaction is taking place and those in which there is a breakdown (TD or BD), the interaction is not yet started, or it is already ended, in line with the UE-HRI coding scheme. Consequently, engagement values are predicted over the RGB image streams from the Pepper robot's head camera. By setting a threshold value thr, such predictions y' are converted into a binary classification of $C = \{\top, \bot\}$ (prediction above or below thr) indicating whether there is engagement or not. The categorical predictions are then compared with values from the annotations in the dataset. The ground truth value is considered to be $y^t_{int} = \top$ if at time t there is an annotation of a *Mono* or *Multi* interaction and there are no annotations of BD or TB in the UE-HRI coding. The ground truth value is $y^t_{int} = \bot$ otherwise.

4.7 Results

With the presented evaluation, I set out to provide evidence that the proposed model can predict engagement through regression on the TOGURO dataset by measuring its accuracy in comparison to the ground-truth annotations, and to assess the generalisation ability on newly encountered situations through the analysis of the UE-HRI data.

To show the ability of the presented framework to map short-term human behavioural features from image sequences into engagement scores, the Mean Squared Error (MSE) prediction loss is computed. It is reported an error on the test set of 0.126 (in the context of the $[0, 1]$ interval of output expected), also shown in Table 4.3.

Table 4.4: Model performances on different conditions of users group composition in terms of MSE test loss and Spearman's Correlation ρ (with smoothing factor $S = 1$ [sec]) of predictions with the ground truth values.

Condition	Test loss (MSE)	Correlation ρ
single-party	0.087	0.758
multi-party	0.136	0.622
adults-only	0.068	0.812
adults-and-minors	0.149	0.563

Additionally, the reported Spearman's rank correlation ρ between the model's predictions and the ground truth values shows that it is consistent with the inter-rater agreement results between the annotators reported in Figure 4.4.2. The results of evaluating the trained model on the four different conditions on the group of users, reported in Table 4.4, show that the model is able to predict the engagement more accurately in the *single-party* and *adult-only* conditions. Comparing these results with the inter-rater agreements for the different conditions, shown in Figure 4.7, it can be observed that the model achieves greater performances in the conditions with more consistent labelling among the different annotators –i.e. *single-party* and *adult-only* where coder 1 and 2 showed generally high correlation and had the highest amount of overlapping videos.

Looking back at section 4.1, a soft real-time operation is seen as a requirement for the applicability of the model. Hence, Table 4.3 reports that the duration of a forward pass on the robot's GPU hardware of 10 consecutive frames (1 sample) through the convolutional module and the recurrent module taking at most $200ms$, allowing real-time estimation of engagement at 5 frames per second.

In order to capture the generalisation capabilities of the trained model, the approach was evaluated for the capability of performing binary classification on the UE-HRI dataset as detailed above in Subsection 4.6.3. Figure 4.9 reports the Receiver Operating Characteristic (ROC) and the Precision-Recall curves obtained by varying the threshold with values in the range $thr \in [0, 1]$ of the binary classification task on the UE-HRI data. The Area Under the Curve ($AUC = 0.89$ in this experiment) reports the probability that the classifier ranks a randomly chosen positive instance $y^t_{int} = \top$ higher than a randomly chosen negative one $y^t_{int} = \perp$, i.e., provides a good assessment of the performance of the model in this completely different dataset.

Given these encouraging quantitative results, some qualitative assessments of exemplary frames with the corresponding computed engagement score are presented in Figure 4.10, Figure 4.11 and Figure 4.12. All figures show examples of the UE-HRI dataset, which was completely absent from the training dataset (Subsection 4.6.1). Figure 4.10 presents two short sequences (roughly 2 seconds apart between frames),

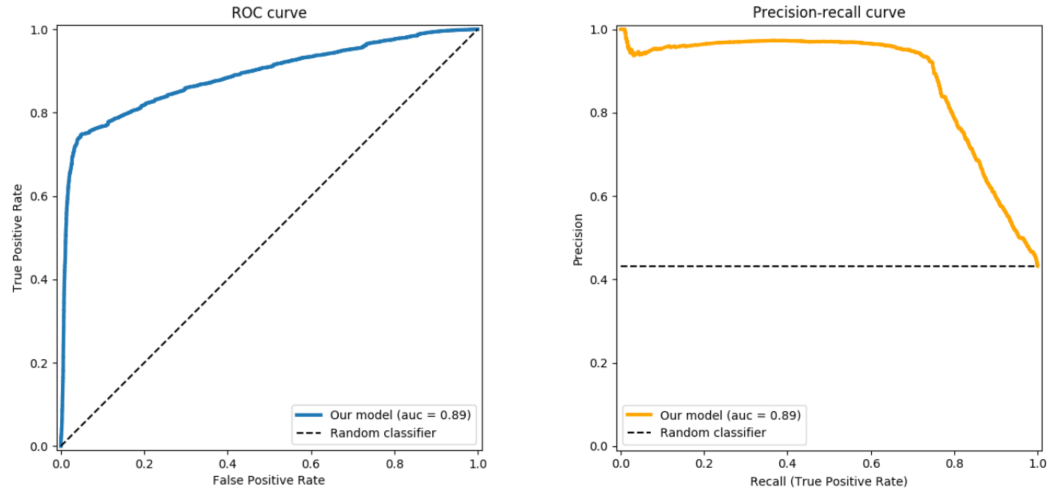
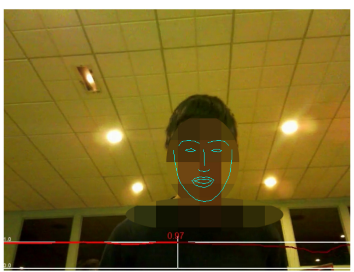
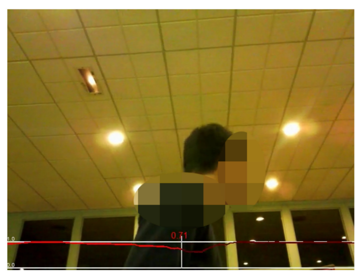


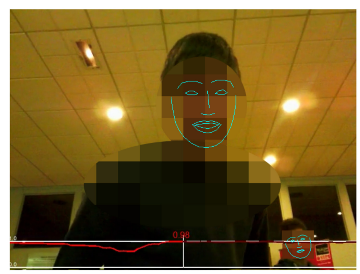
Figure 4.9: ROC curve (on the left) and Precision-Recall curve (on the right) generated using our trained model as a classifier of the interaction sessions for the UE-HRI dataset.



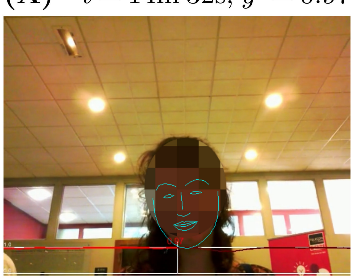
(A) $t = 14\text{m } 32\text{s}$, $y' = 0.97$



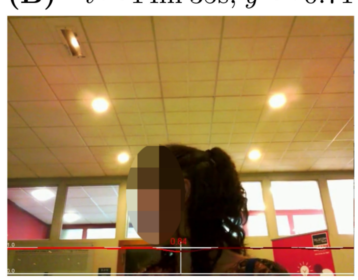
(B) $t = 14\text{m } 35\text{s}$, $y' = 0.71$



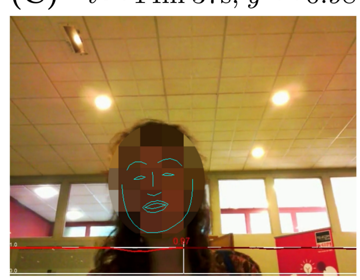
(C) $t = 14\text{m } 37\text{s}$, $y' = 0.98$



(D) $t = 5\text{m } 56\text{s}$, $y' = 0.97$

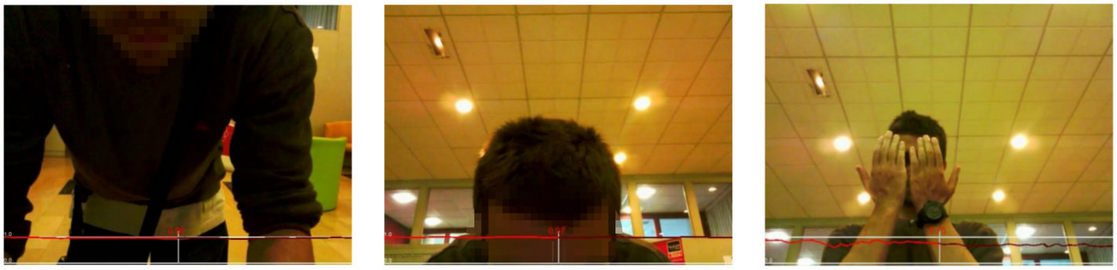


(E) $t = 5\text{m } 58\text{s}$, $y' = 0.84$



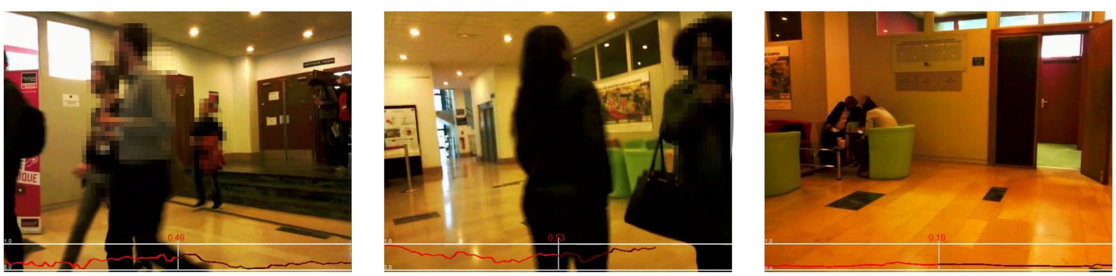
(F) $t = 5\text{m } 59\text{s}$, $y' = 0.97$

Figure 4.10: UE-HRI dataset: two sequences of short timescale sequential frames showing how the temporal diverting of attention is reflected in the model predicting a lower engagement value. The red plot shows the predicted engagement values over the frame sequences, with the prediction y' at the frame shown in the picture at time t being in the centre, past predictions on the left and future predictions on the right. Faces from the original dataset are blurred for anonymisation; face landmarks are added in post-processing to indicate face orientation. Permission for re-use of the images was obtained from the copyright holder [12].



(A) $y' = 0.96$ **(B)** $y' = 0.97$ **(C)** $y' = 0.75$

Figure 4.11: UE-HRI dataset: examples of correct prediction of high engagement ($y' >= 0.75$) in situations difficult to understand using standard face description features. The red plot shows the predicted engagement values over the frame sequences with the prediction y' at the frame shown in the picture being in the centre, past predictions on the left and future predictions on the right. Faces from the original dataset are blurred for anonymisation. Permission for re-use of the images was obtained from the copyright holder.



(A) $y' = 0.58$ **(B)** $y' = 0.53$ **(C)** $y' = 0.19$

Figure 4.12: UE-HRI dataset: examples of correct low/medium engagement prediction ($y' <= 0.6$) in cases in which the people were not actually engaging with the robot. The red plot shows the predicted engagement values over the frame sequences with the prediction y' at the frame shown in the picture being in the centre, past predictions on the left and future predictions on the right. Faces from the original dataset are blurred for anonymisation. Permission for re-use of the images was obtained from the copyright holder.

showcasing short-term diversion of attention of subjects resulting in a temporarily lower engagement score but not leading to a very low engagement level. Figure 4.11 exemplifies that the regression model can cope well with perception challenges which would forgo a correct assessment just using gaze or facial feature analysis. While one could, in this context, argue that the model has simply learned to detect people, Figure 4.12 is providing three examples from different videos of the UE-HRI dataset with people present in the vicinity of the robot, but not engaging with it. The engagement scores in these examples are significantly lower across all frames.

These qualitative reflections are evidently supported by the quantitative analysis of both datasets, indicating that the trained model is broadly applicable and can serve as a useful tool to the HRI community with its modest computational requirements and high response speed in assessing users' engagement from a robot's point of view.

4.8 Discussion and Conclusion

This chapter has motivated, developed and validated a novel and ready-to-use computational model to assess engagement from a robot's perspective. The results presented in the previous sections lead to the conclusion that:

i a moderate to strong inter-rater agreement (see Table 4.2) in measuring engagement on $[0, 1]$ interval indicates that humans can reasonably and reliably assess the holistic engagement from a robot's point of view solely from video, validating the **Hypothesis 2** formulated in Section 1.1;

ii a two-stage deep-learning architecture as presented in Figure 4.8 trained from our TOGURO dataset is a suitable computational regression model to capture the inherent human interpretation of engagement provided by the annotators;

iii the trained model is generic enough to be successfully applied in a completely different scenario, here the UE-HRI dataset, showing the applicability of the model also in different environments, on a different robot with a different camera, and with different tasks and people. The area under the Receiver-Operator Curve (ROC) of 0.89 and the Precision-Recall curve in Figure 4.9 provide evidence that the proposed regression model can serve as a strong discriminator to identify situations of loss of engagement (TD or BD in the UE-HRI coding scheme). This fulfills this thesis's **Objective 3**.

Such conclusions confirm the idea that the human holistic assessment of an abstract quantity, like engagement, can be utilised as a coherent metric for learning a prediction model of that same quantity. This idea was never tested before specifically on engagement, however, it follows approaches from Tanaka et al. [91] where "quality of the interaction" was similarly measured with a continuous coders' assessment. Therefore, the presented work provides further evidence of the suitability of the methodology. While a cue-centric model based on specific perceptual features, such as gaze, can be more easily interpretable, it can miss out on important events that are not explained by the chosen features. A model learned from raw data, like the one presented in this chapter, can instead learn to recognise the important features to take into account for the assessment. The approach of learning representations of raw data with DL models, is becoming standard practice in social robotics solutions thanks to their ability to learn good quality representations of human-related attributes, e.g. [73, 13, 79].

The hypothesis behind the model choices is that the learned model does not solely discriminate person and/or face presence, but that the temporal aspects of the human behaviour observable in the video are captured by the LSTM layer in our architecture well enough to successfully deal with these situations. The correlation values between the model predictions and the ground truth value in Table 4.3 and Table 4.4 suggest that the predictions are in line with the coders' assessment, even when averaging at a short timescale like 1 second. This result is important because it shows that the model can be used to immediately identify moments of decreased engagement and plan to recover from them before users completely disengage with the robot. By embedding the results of this chapter into the wider context of the aims of this thesis –i.e. the online adaptation of the robot behaviour during human-robot interactions–, it is not difficult to imagine how the engagement model's estimations can be used as a metric which assigns a scalar value to each action executed by the robot. In this way, such values are provided to represent the goodness of the action in terms of maintaining users' engagement during the interaction. With this general idea in mind, the next step to producing a general framework that can adapt robot behaviours online, is to use Reinforcement Learning (RL) techniques by integrating our engagement model predictions as part of the reward function given to the robot. As typically done in RL problems, the reward function provides only a scalar value at each time step for the agent to learn how to associate good actions in the different states encountered during exploration.

4.9 Summary

This chapter presented a regression model that can assess the user's group engagement value in real-time during human-robot interactions from the robot point-of-view. The approach proposed consists of a learned model trained from videos collected during spontaneous interactions in a public museum and annotated with holistic human assessments of engagement.

Unlike previous engagement detection approaches, the regression model does not rely on specific features provided by other lower-level computational modules like gaze or face orientation, but rather uses raw video frames. Following the same principle of not relying on specific cues, the coders that annotated the videos did not attempt to identify specific user behaviours or features but were instructed to provide their intuitive evaluation of user engagement as if they were delivering the tour in place of the robot.

The results showed that, despite trying to sidestep as much as possible reliance on explanations for the displayed users' engagement, humans are able to provide a ground truth with medium to high agreement correlation and the DL model proposed is able to learn from these holistic assessments. The engagement scalar assessments produced by the learned model in the $[0, 1]$ range can be further assumed to represent evaluations of the robot's actions. In cultural experiences like guided tours in museums where the desired outcomes of the interactions for users are learning and entertainment, an essential feature of the guide is the ability to deliver engaging interactions. This idea motivates the work in the following chapter, which uses the engagement assessments of the regression model as rewards to optimise the robot behavioural policy.

CHAPTER 5

In-Situ Behavioural Adaptation

This chapter presents a framework for the online adaptation of autonomous social robot behaviours, where the behaviour is initially manually specified and improves over time after each user interaction. The adaptation algorithm allows the robot to improve its effectiveness at interacting with people by learning to select actions expected to elicit high user engagement values. In order to achieve such aim, this chapter puts together the robotics framework enabling the long-term deployment in the museum, taking into account the lessons learned during it, described in Chapter 3 and the engagement model described in Chapter 4. The following sections describe the settings of the scenario from which the data comes, the system design to allow the model to learn online from real-world interactions, and the evaluation of such a system in the ongoing long-term museum deployment.

5.1 Introduction

Maintaining user engagement alive during interactions with a robot is essential for social robotics technologies. In Chapter 3 it was shown how user engagement with Lindsey in the museum is easy to start, but quickly decreases after the first minutes of interaction. The situation where a robot is initially very engaging for users for the sole virtue of being a novel technology is a common phenomenon in social robotics applications and is generally referred to as *novelty effect*. In order to keep such initial interest alive for longer, robot behaviours need to be interactive and consider the interaction's contextual factors.

The approach followed in this chapter is to allow the robot to modify its behaviour and to learn from experience how the execution of different actions affects the users' state and the ongoing interaction. Reinforcement Learning (RL) techniques are a special case of Machine Learning (ML) algorithms that deals precisely with such scenarios in which the "goodness" of the actions an agent can take is not known in advance, and exploration is required. The goal of the optimisation performed by RL methods is finding the best sequence of actions to maximise a particular objective, manifested through rewards. However difficult this may seem for scenarios where the

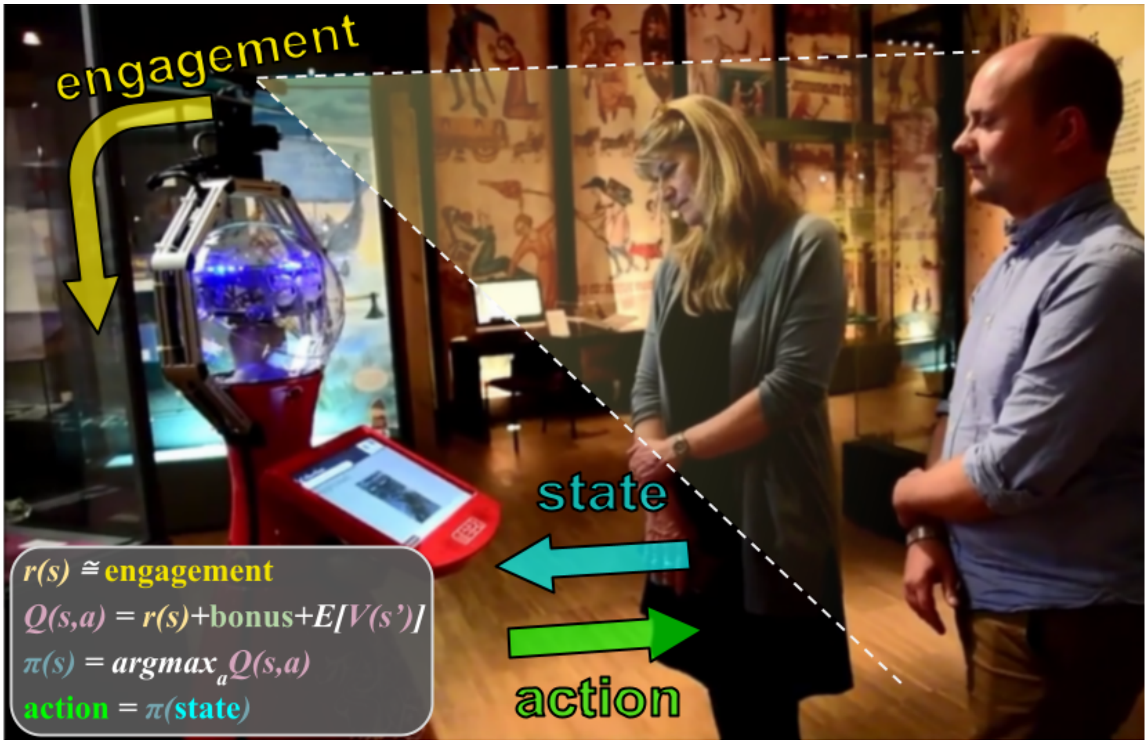


Figure 5.1: A picture of Lindsey in the archaeological gallery of The Collection museum during a tour guide interaction. The robot learning framework detects user engagement online (from the work described in Chapter 4) and optimises its interaction policy for maximising such overall engagement using Reinforcement Learning (the work described in this Chapter).

goal is well defined, it becomes even more challenging for social scenarios where the objective is expressed by the users' affective state (assuming that we are maximising positive affective states during the interactions) and can only be estimated from sensors or proxy variables (like the duration of the interaction) since it is not a direclty accessible measure.

The work presented here, and published in [28], solves this problem by enabling robots to estimate the users' engagement during the interaction and allowing them to learn, through RL, the actions to maximise such engagement. This framework is built on the previous work of Chapter 3, and published in [26], which described and analysed the long-term deployment, which is still ongoing with the present work, of a robot in a public museum where it serves as a tour guide to the visitors. For detecting the users' engagement, the learned regression model proposed in Chapter 4 and published in [27] is used, providing a single scalar engagement from standard video streams obtained from the point of view of the interacting robot. To illustrate the overall idea of this chapter Figure 3.1 shows Lindsey the robot in the museum while interacting with the users and depicts the concept for the proposed learning framework.

By learning in a long-term scenario using only the users' engagement as guidance, this work provides a proof-of-concept for a framework to enable behavioural adaptation in social robots. Given that the learning is guided by the users' state, manifested by their expressed engagement, this will allow the robot to adapt to the users' preferences, which cannot be known or programmed in advance.

5.2 Related Work

This chapter presents a system that can learn from interactions to perform better engagements with the users during long-term deployments in a public space. Here, a review of work that has addressed the problems of having robots interacting autonomously in public spaces is performed. In particular, the focus is on the methods that can be used in such context to influence the users' emotional state or their engagement and that allow the robot behaviour to adapt accordingly.

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 [64, 15]. However, in public spaces, such as museums, the importance of considering the humans in the environment goes beyond simply exploiting their help to perform the robot's tasks; it is necessary that the robot can understand and adapt to the users [52]. In a long-term deployment within the STRANDS project, Hawes et al. [40] report the need to have a better understanding of human activities and, in that context, Hanheide et al. [39] propose a spatio-temporal model to learn when, where and how users interacted with the robot info-terminal during a long-term deployment. They found they could improve the efficiency and usefulness of the system by proposing the right content at the right time and place. In a survey on long-term interaction between users and robots Leite et al. [55] raises the issue that memory and adaptation remain nearly unexplored in the field, proposing the use of "strategic behaviours" that the robot should enact in order to preserve or improve the relationship with its users.

The following describes previous works that have shown examples of how it is possible to influence the users' engagement level within the interactions and, subsequently, to incorporate the users' feedback to adapt the robot behaviours accordingly.

5.2.1 Influencing the Users' Engagement with a Technology

In social human-robot interactions, where the goal of the robot operations is to engage with the users, the ability to positively influence the users' internal state is an essential aspect. The sole presence of robots, particularly if novel, is sufficient for higher engagement in learning activities, as shown by Baxter et al. [9]. However, this is not sustained over time as shown by previous work, e.g. [56, 47], and confirmed by the previous work [26] described in Chapter 3.

Despite the fact that engagement is usually soon lost in human-robot interactions, other works have shown that the appropriate use of technology can effectively improve users' emotional state and increase their engagement in the context of social interactions and learning activities. Sidner et al. [87] explored how the use of gazing and gestures positively affects the user perception of the robot, increasing their engagement. Similarly, Holroyd [43] defines policies to increase user engagement and shows that the robot equipped with these policies is perceived to be more human-like, to behave more fluently and that users reciprocate more robot cues. Also, between works that focus on the use of technology for learning and therapy, developing engaging strategies is important. For example, in the five years of deployment of four robots focused on interactivity and education, Nourbakhsh et al. [69] have learned that long and non-interactive presentations are guaranteed to drive the audience away. Standen et al. [89] shows that engagement increases in learning activities for people with Intellectual Disabilities, when the activity itself is tailored to their personal needs and emotional states. Balasuriya et al. [5] also found that playing games with a Cozmo robot displaying human-like behaviours could increase their engagement in the activity. Similarly, [85] found an increase in engagement when deploying NAO robot interventions in therapy sessions.

Taken together, the literature shows that robots can be an suitable tool for improving and maintaining the user's engagement alive in various different activities. However, in order to be effective and retain the attention after the initial novelty effect the robot needs to be programmed appropriately, for example by deploying appropriate gazing strategies or by personalising the interaction to the specific user. These results motivate the pursuit of better strategies for the robot behaviour,since it has been ascertained that the current "static" behaviours cannot keep the users' engagement alive for long (i.e. see Chapter 3). In particular, the work presented here attempts to tailor the robot guided tour to the specific engagement state of the user, since personalisation to the specific users is not possible in a museum environment with an always-changing flow of visitors.

5.2.2 Behavioural Adaptation in Social Settings

Approaches in the literature for programming appropriate robot behaviours that are able to sustain social interactions can be divided into those that seek to personalise the actions to the specific users and those that instead try to learn behaviour based on the immediate state of the users.

Within the first group of works, Kubota et al. [51] propose a method for synthesising robot behaviours from high-level specifications given by non-expert people, allowing clinicians to adapt the therapy to their patients. In this approach, the robot's behaviour is manually specified and not automatically learned based on the user characteristics. Similarly, Ligthart et al. [60] provide a memory-based personalisation strategy of a social robot interacting with children over multiple sessions being able to keep the children's interest alive over time. In order to learn the best policy tailored to each individual characteristics, Andriella et al. [2] propose the use of a persona-based simulation to categorise users based on their capabilities to learn, using RL, robot policies tailored to their specific needs. Similarly, in [1] a framework is proposed to learn assistive behaviours by leveraging therapists' expertise and demonstrations.

Note that in public spaces, like museums or shopping centres, personalisation to the individual users is not a feasible option given that there is usually a constant flow of new people interacting with the robot, and it may not even be an appropriate approach for privacy concerns. However, in such long-term scenarios, adaptation is still necessary [55], and it can be performed by exploiting the users' state during the interaction, rather than their individual characteristics. In Tanevska et al. [92], the robot behaviour adapts based on the amount of stimuli received during the interaction, however, the behaviour was driven by a cognitive model and not learned. Using a learning approach, Senft et al. [83] propose a method for learning an autonomous robot policy where the robot proposes actions that can be accepted or rejected by a supervisor, progressively learning from such guidance. The Minerva robot [93], which traversed more than 44 km and interacted with more than 50k people in a public museum, was able to display mood (i.e. happy or angry) and used an RL approach to learn the best actions to engage visitors based on their vicinity to the robot. Recent works aimed at learning these social behaviours typically use (Deep) RL techniques to exploit the real-world interaction experiences a robot can collect. Mcquillin et al. [63] propose a Deep RL approach to learn a robot policy for approaching humans by exploring the use of an explicit reward, i.e. positive and negative keywords in user's speech, and an implicit one, computed from the valence score of participants. Similarly, Qureshi et al. [73, 74] proposed end-to-end models for teaching a robot the most appropriate action for approaching humans and starting an interaction. Successful/unsuccessful handshakes triggered the reward signal. Lathuilière et al. [53] uses Deep RL to learn a gaze policy from an intrinsic reward function based on the audiovisual position of people with respect to the robot camera field of view. Gao et al. [35] learn a robot policy for approaching groups of people by maximising a group formation score and minimising the displacement of other participants in the group when the robot approaches. In a museum context, Meng et al. [65] propose an RL approach where they gather the users' engagement during group interactions with an interactive sculpture as the reward for learning engaging interactive behaviours. Although, this approach was not deployed on a social robot, it shows an example of how continuous engagement estimation can be used as an extrinsic reward for learning.

In the work presented here, taking inspiration from the above literature a RL model is trained using an extrinsic reward based on the users' state during the interaction to maximise user engagement. Similarly to [65], the reward is made of a continuous supervisory signal representing the user engagement; however, here the signal is generated by the engagement regression model described in Chapter 4. Moreover, the action space of Lindsey the robot is a discrete one where the learning can lead the robot to give more or less detailed information about each item and to change the order with which they are described in each tours.

5.3 Preliminaries

This section introduces the methodologies the presented adaption framework is based on. The framework uses RL for finding the best actions the robot should execute at any moment during the tour based on the user engagement level, therefore, the basic components of this learning paradigm are reviewed below. Given the constraint of learning in a real-world HRI deployment with a limited amount of interactions possible, exploration is limited and an efficient learning strategy needs to be implemented. Therefore, approaches inspired by the multi-armed bandits field are presented as they provide valuable tools for optimal exploration.

5.3.1 Reinforcement Learning

RL is the study of intelligent decision making, i.e. it is concerned with the problem of how an agent should take actions in the environment in which it is embedded. The objective of reinforcement learning algorithms is to find a model of the agent behaviour, referred to as a policy π, that select the action the agent ought to take in each state so that the reward accumulated over time is maximised.

Mathematically, the RL problem can be described by the Markov Decision Processes (MDP) formalism which frames the problem of sequential decision making in which the actions taken by the agent influence the immediate reward and the subsequent state of the environment (which includes everything outside the agent). Figure 5.2 shows the reinforcement learning loop described by the MDP formalism. At any time step t the agent observes the state of the environment $s_t \in S$, possibly receives a reward r_t for being in that state and executes an action $a_t \in A$. The rewards are received by the agent upon executing an action based on the state of the environment afterwards and are generated by the environment itself. In some special cases, like intrisic motivation approaches [7], the rewards are automatically generated by the agent. If known, the dynamics of the MDP are specified through a model of the environment. That is, a function p that for each $s \in S$ and $a \in A$ gives the probability of the next state $s' \in S$ and reward $r \in \mathbb{R}$, i.e.

$$p(s', r|s, a) \doteq Pr\{s_t = s', r_t = r|s_{t-1} = s, a_{t-1} = a\}.$$

Differently than supervised learning problems, where the correct output is provided in the training data, in reinforcement learning the model of the environment is not known (beyond toy problems) meaning that the agent does not know in advance the "goodness" of the actions it can select and it has to explore them by first acting in the environment. Even after exploration, the agent is not given then "correct" action to execute but only has an evaluation of it, hence it cannot know if different

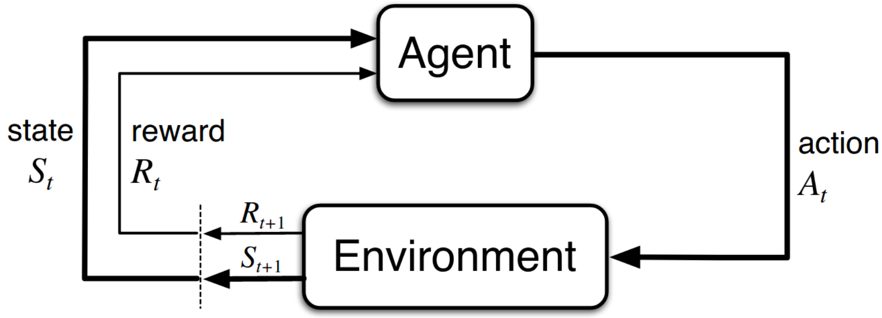


Figure 5.2: The interactions between an agent and its enviornment in Reinforcement Learning problems. Figure from Sutton and Barto [90].

actions are better or worse until full exploration is performed. However, in most useful problems, full exploration of the state-action space is unfeasible and we must rely on approximate values. The dilemma of whether to rely on the potentially sub-optimal estimated values to choose the next best action, or to explore more to get a better approximation for the future is referred as *exploration-exploitation trade-off*. A simple yet often effective exploration mechanism is ϵ-greedy, where $0 < \epsilon < 1$ and the action is choosen randomly with probability ϵ or else by exploiting of the current best action.

How good is being in a particular state s, or being in a s after executing an action a, is identified through *value functions* which specify the amount of reward the agent is expected to accumulate over time starting from that state. $V_\pi(s)$ is the name used to refer to value functions over the state space while $Q_\pi(s,a)$ specifies the value based on executing an action in a particular state. Value functions vary when choosing different policies given that the expected cumulative reward depends on the sequence of future action choices. By maximising the value function, one attempts to find the *optimal policy* π_* which is the one that has the highest expected returns over the other possible policies, i.e. if $V_{\pi_*}(s) \geq V_{\pi'}(s)$ for all $s \in S$ and for any possible choice of π'. The value functions of the optimal policy, called *optimal value functions*, are defined as

$$V_*(s) \doteq \max_\pi V_\pi(s), \text{ for all } s \in S; \text{ and}$$

$$Q_*(s,a) \doteq \max_\pi Q_\pi(s,a), \text{ for all } s \in S \text{ and } a \in A.$$

5.3.2 Dynamic Programming

Dynamic programming is a family of algorithms that can be used to find the optimal policy in a MDP given that a perfect model of the environment exists. The general principle of dynamic programming approaches is the subdivision of the problem into smaller and easier subproblems. The solution to the original problem is simply found by combining the solutions to the subproblems. Various algorithms using dynamic programming have been proposed for solving the MDP, however all of them share the same two building blocks of *value update* and *policy update*. Both updates are fomulated recursively where new estimations of the optimal policy and optimal value function are computed from a previous estimate. The value update step is derived directly as the Bellman equation [11], which is the necessary condition to satisfy the optimality of the value function and therefore computed as

$$V_{i+1}(s) := \max_a \sum_{s',r} p(s',r|s,a)\left[r + \gamma V_i(s')\right],$$

or

$$Q_{i+1}(s,a) := \sum_{s',r} p(s',r|s,a)\left[r + \gamma \max_{a'} Q_i(s',a')\right],$$

where i is the step number.

The recursive operation, if applied enough times, will allow the value function to converge to the fixed point $V_{i+1}(s) = \max_a \sum_{s',r} p(s',r|s,a)\,[r + \gamma V_i(s')] = V_*(s)$. Once the optimal value function V_* is known, the optimal policy is easily obtained:

$$\pi_*(s) = \arg\max_a \sum_{s',r} p(s',r|s,a)\left[r + \gamma V_*(s')\right].$$

The Value Iteration (VI) algorithm follows exactly this paradigm where the Bellman's operator is applied repeatedly on an initial value function until the values converge, hence the name. Theoretically the algorithm requires an inifite number of steps to converge but, in practise, the algorithm can be stopped when the difference betwen the older and the newer values are small enough. Note that in VI the policy is only recovered at the end of the iterative process by "greedily" chosing the actions with the highest expected value given the state.

5.3.3 Multi-armed Bandits

The multi-armed bandits problem is a simplified version of the reinforcement learning problem where the agent has to act always in the same state, rather than in an environment that changes as the agent acts in it. In fact the multi-armed bandit literature studies what is the best way to act when you are repeatedly faced with the same choice of actions and the reward received after each action is drawn from a distribution (which can be non-stationary) dependant on the chosen action. Given that the environment is stateless, the only learnable part of the environment is the reward model but exploration is still required as this is not known in advance. Multi-armed algorithms keep estimates about rewards through the action value function $Q(a)$ which, in the simplest case, can be implemented as a simple average of the rewards received for an action $Q_t \doteq \frac{r_1+r_2+\ldots+r_{t-1}}{t-1}$, with t the number of times the action has been executed. The *greedy* action is simply found as

$$a' \doteq \arg\max_a Q(a)$$

The simplified nature of the problem allows to study strategies for balancing exploration and exploitation in a clearer form than in the full reinforcement learning setup. However, good balancing strategies for the multi-armed bandit problem are usually good strategies for the RL problem as well.

A simple yet effective method for balancing exploitation with exploration is what is referred to *optimistic* approach. Following this paradigm means that the agent, when missing a reasonable estimate of the model for a specific state, it assumes the best possible outcome. If the model is correct, you have *no regrets* (exploitation); otherwise, you have effectively learned something new about the world (exploration). This allows to quickly explore actions that have been rarely chosen, while exploiting those that have had good returns so far. Practically, one way of implementing this optimistic behaviour is to select actions according to

$$a_t \doteq \arg\max_a \left[Q_t(a) + c \cdot \sqrt{\frac{\ln t}{N_t(a)}} \right],$$

where $\ln t$ is the natural logarithm of t, $N_t(a)$ counts the number of times action a has been selected and $c > 0$ is a factor controlling the amount of exploration.

This general principle gives an almost optimal solution for the stochastic multi-armed bandit problem [17] and episodic [3] and ergodic [46] RL problems.

5.3.4 Upper-Confidence Bound Value Iteration

The Upper Confidence Bound Value Iteration (UCBVI), proposed by Azar et al [3], is an algorithm for Reinforcement Learning problems which frames the VI algorithm within the paradigm of optimistic exploration. This extension of the Value Iteration algorithm guarantees that the value function found is with *high probability* an upper confidence bound on the optimal value function. It assumes that the problem is episodic, i.e. that the agent interacts with the environment in a series of episodes each with an initial and final state, and it does not require the model of the environment to be known in advance, but it is learned as the agent explores.

The algorithm is composed of 2 main steps: 1) action-values estimation, which computes the optimistic Q values based on the experience collected so far; and 2) policy execution, which uses the greedy policy from the Q values to act in the environment and collects experience data. The high-level pseudocode is shown in Algorithm 1.

The optimistic approach is implemented by summing a `bonus` to the action values. Similarly to the optimistic multi-armed bandit algorithm the bonus factor typically depends on the number of times the action has been explored, it's higher for poorly explored actions and descreses as more exploration is performed.

Algorithm 1: UCBVI

```
Data: S, A, H
𝔻 ← ∅;
foreach episode do
    // action-values estimation
    Estimate r̂_h(s) and p̂_h(s' | s, a) from 𝔻;
    for h = H, ..., 1 do
        for (s, a) ∈ S × A do
            b_h(s, a) ← bonus(h, s, a);
            Q_h(s, a) ← r̂_h(s) + b_h(s, a) + E_{s'~p̂_h(·|s,a)}[V_{h+1}(s')];
            V_h(s) ← max_{a∈A} Q_h(s, a);
        end
    end
    // policy execution
    Observe s_1;
    for h = 1, ..., H do
        Execute a_h = arg max_{a∈A} Q_h(s_h, a);
        Observe r_h and s_{h+1};
        Add (s_h, a_h, r_h, s_{h+1}) to 𝔻
    end
end
```

5.4 The Behavioural Adaptation Framework

This section outlines a Reinforcement Learning (RL) framework that adapts the high-level behavioural policy of a robot (which is initially manually specified) with the goal of improving the quality of the interactions by integrating the engagement of the people interacting with it. The framework is based on the robotic software architecture described in Chapter 3; in particular, the static guided tours plans are taken as a starting point for the robot policy, and the users' engagement model shown in Chapter 4 for generating a learning signal. The Figure 5.3 shows the classical RL loop augmented with the engagement model and with indication at how the various components have been implemented in this thesis. The overall system allows the robot to explore different actions and learn online during the interactions with the users during the deployment.

5.4.1 States & Actions Specification

The states and actions that are considered for behavioural adaptation are those from the guided tours behaviour which has been selected for its complexity and duration, hence allowing a more challenging setting for the learning framework. The robot's initial behaviours are defined as conditional plans using the Petri Net Plans (PNP) formalism. The conditional plans are specified as sequences of actions, conditions,

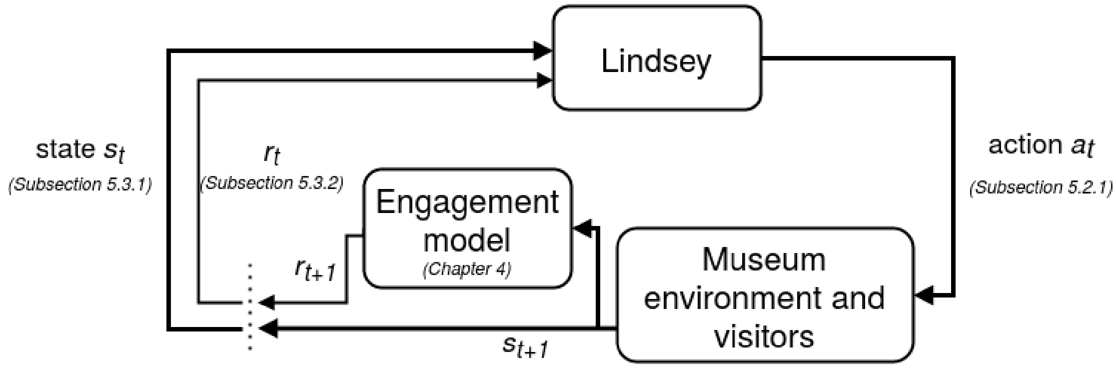


Figure 5.3: A re-adaptation of Figure 5.2 which grounds all the various components of the RL loop in the specific scenario of this work.

and *execution rules* for monitoring the execution of the actions. The plans are inherently hierarchical given that each plan can be called as an action by other plans. More detailed information about the specification of the behaviour using PNP can be found in Subsection 3.4.2. For the purpose of allowing adaptation of the behaviours, the plan at the highest level of the guided tour behaviour is replaced by a policy, whereas the sub-plans and the atomic actions (i.e. the low level implementation of actions) are left untouched. In the remainder of this chapter the word "action" will be used to refer to both atomic actions and sub-plans.

Table 5.1: Action space with constraints over possible action successors. The successors in grey are conditional on that action not being executed already in the tour and those underlined are only valid for the `death` tour.

Id	Action	Action Successors
0	`doNothing`	0,1
1	`describeTour`	2,3,4,5,6,7
2	`gotoExhibit_1`	8
3	`gotoExhibit_2`	8
4	`gotoExhibit_3`	8
5	`gotoExhibit_4`	8
6	`gotoExhibit_5`	8
7	`gotoExhibit_6`	8
8	`describeExhibit`	2,3,4,5,6,7,9,10
9	`describeMoreExhibit`	2,3,4,5,6,7,10
10	`endTour`	0

Given the nature of the robot behaviours, i.e. navigating to and from different places in the museum and describing items with different amount of words, the duration of actions can vary from seconds to minutes and is not known in advance. Any action is also interruptible and it can potentially terminate the entire tour before it reaches the end, for example, in case of failures or the user actively stopping the tour. Moreover, the guided tours behaviours need to be coherent, requiring certain actions to be performed exclusively in a specific sequence. For example, it would be wrong to start describing an exhibit's item before going to its location or describing the theme tour only at the end of the tour. The `gotoExhibit_*` actions cannot be executed more than once because it wouldn't make sense bringing people to the same item in the museum multiple times during a tour. In order to enforce these constraints at execution time, it was necessary to created subsets of all the admissible actions $C_s \subset A \; \forall \; s \in S$ from which the policy can pick a successive action based on the current state. The action space with the associated successive action constraints is showed in Table 5.1. Note that differently than the initial definition of the guided tours plans described in Subsection 3.4.4, now the robot is allowed to visit the sequence of exhibitors in each tour in different orders and it does not ask the users if they want to receive more information, but rather the policy can choose whether to give additional informations or more forward with the tour. The size of the action set is $|A| = 11$, while the size of the action subsets is $1 \leq |C_s| \leq 8$.

The state vector used for learning is composed as follows

$$s = \langle v_1, v_2, v_3, v_4, v_5, v_6, n, a_p, e, t \rangle$$

where $v_k \in \{0, 1\}$ for $k = 1, ..., 6$ indicates whether the k^{th} item in the tour has been visited, $n \in \{\texttt{death}, \texttt{tools}, \texttt{religion}, \texttt{art}\}$ is the name of the tour, $a_p \in A$ is the previous action, $e \in \{\texttt{LOW}, \texttt{MEDIUM}, \texttt{HIGH}\}$ is the average engagement level during the execution of a_p, and $t \in \{\texttt{none}, \texttt{ended}, \texttt{stopped}, \texttt{abandoned}\}$ indicates the terminal state. Putting all the fields together, size of the state space is $|S| = 33\,792$.

5.4.2 Engagement Model

The engagement of users who interact with Lindsey at the museum is detected using the engagement model presented in Chapter 4, which runs in real-time on the robot's GPU. The model consists of an end-to-end regression model, providing a holistic engagement measure in the interval $[0, 1]$ as output to each sequence of image frames taken from the robot's camera stream. It was validated on HRI data sets, i.e. the TOGURO dataset described in Section 4.4 and the UE-HRI dataset [12], as an accurate measure of engagement. Figure 5.4 depicts 3 frames from a scene during a guided tour showing the continuous engagement signal provided by the model. In the adaptation framework, the user engagement signal is used as an evaluation metric of the robot's behaviours to guide learning. Therefore, in this RL setting the reward observed at each state is derived from the engagement collected during the execution of the last action. The engagement scalar values are provided by the model at a frequency of about 1 Hz, they are then averaged over the entire duration of the action and discretised in the variables `LOW`, for values in $(0, \frac{1}{3}]$, `MEDIUM`, for

Figure 5.4: Examples of continuous engagement predictions from the robot head camera using the model from Chapter 4. Engagement prediction value is `LOW` in the top frame (the girl is looking elsewhere), `MEDIUM` in the centre frame (man is taking a picture of the robot) and `HIGH` in the bottom frame (man looking at robot screen). Written consent was obtained from the users for taking and reusing these pictures.

values in $(\frac{1}{3}, \frac{2}{3}]$, or HIGH, for values in $(\frac{2}{3}, 1]$. Finally, the reward function, given the state s, is

$$r(s) \sim \begin{cases} \mathcal{U}(0, \frac{1}{3}) & \text{if } e = \texttt{LOW}, \\ \mathcal{U}(\frac{1}{3}, \frac{2}{3}) & \text{if } e = \texttt{MEDIUM}, \\ \mathcal{U}(\frac{2}{3}, 1) & \text{if } e = \texttt{HIGH}. \end{cases}$$

Such scalar value for engagement is given as the reward to guide the learned policy toward actions that are expected to elicit higher future users' engagement. Given that $r(s) > 0 \; \forall s \in S$ there is an additional implicit effect of favouring longer tours. However, longer tours with lower detected users engagement are not necessarily better than highly engaged short tours. Therefore, the RL algorithm aims to find the best balance between those two conditions.

5.4.3 Behaviour Adaptation with UCBVI

Adaptation of the robot behaviour is performed using an RL approach where the policy starts as a handcrafted plan –i.e. the static guided tour defined manually by the educators of the museum as described in Subsection 3.4.4– and it gets updated over time as more states and actions are explored. The goal of such behavioural adaptation is to validate a system that can automatically learn engaging behaviours for social robots from interactions with users, without requiring explicit feedback from the users or a handcrafted reward function. For this reason, the holistic users' engagement assessment has been chosen as the metric for driving learning and the guided tour as a benchmarking scenario which allows to have a structured social interaction so that different actions can be explored and evaluated against each other by the robot. The structure of the robot behaviour is given by the constraints of the guided tours as described in Subsection 3.4.4 and Subsection 5.4.1.

A model-based approach is proposed to model the dynamics of parts of the environment that are already known, for example the fact that transitions from an *art* tour state to a *death* tour state is not possible. This prevents the robot from performing actions that are not coherent and allows to increase the sample efficiency during learning, given that the constrained parts of the state-actions space do not need being explored. Moreover, in the ongoing deployment, some parts of the state-action space have been explored very well –namely the transitions in the original static tour that have been encountered thousands of times in more than one year of operation– while other areas have never been observed. This means that when estimating the model for the environment, very accurate probabilities are obtained for the already explored areas while maximum uncertainty remains elsewhere. Given that this model is only partly known in advance, the transition probability function $\hat{p}(s'|s, a)$ is estimated from the real data collected during the deployment, while the reward function $r(s)$ is already given as the users' engagement, as described in Subsection 5.4.2. Exploring all the possible state-action pairs is unfeasible in this scenario, both because it would take too long in the real wold deployment, and beacause some states are not directly accessible by the robot as they are manifested also as a result of the users behaviour. Therefore, an efficient strategy is needed to deal with the *exploration-exploitation* problem, i.e. deciding at any moment whether to explore new actions or keep executing the ones that have returned the highest rewards so far.

The setup of this work can be framed as an episodic problem, where every guided tour given by the robot is an episode. In order to improve the robot policy in this episodic scenario, the Upper Confidence Bound Value Iteration (UCBVI) algorithm [3] is used to efficiently balance the exploration of novel state-action pairs with the exploitation of already explored ones. In our setting, this is particularly important because a part of the state-action space was explored very well during the initial deployment with the static policy. UCBVI gives a natural way of incorporating previously accumulated data without biasing the policy to the point that it doesn't explore new actions. Algorithm 2 outlines our implementation of the algorithm. As in standard *model based* Value Iteration, the algorithm is composed of 3 phases. First, the value function is computed from the model and an initial policy is generated; then, the policy is used for acting in the environment while collecting a new episode; finally, the model is improved based on the updated dataset of episodes. The loop repeats, in our case, for every new guided tour that the users request.

UCBVI favours the exploration of novel state-action pairs thanks to a bonus function that increases the Q-value for pairs that were scarcely visited. The bonus decreases toward zero as more data is obtained and the Q-value tends to $\hat{Q}$, i.e. the value function without bonus which after enough exploration should represent the actual estimate of the state-action value. Algorithm 3 shows the bonus, inspired by the Chernoff-Hoeffding's bonus proposed in [3]. The speed with which the bonus tends to zero is controlled by the parameter σ, which can be set based on the settings of the problem. For this work, it was set to $\sigma = 20$ in order to allow a relatively fast exploration of new state-action pairs in our real-world scenario, where obtaining new episodes is difficult. However, it can be set to a lower value in situations where performing exploration is more accessible, like in simulations, for example.

In our implementation, it is essential to notice that the Q-function defined favours the exploration of poorly explored areas over other areas that have never been observed. The effect is that the policy sticks to choosing the same, poorly explored actions over consecutive episodes, starting to explore a new one only after a certain number of visitations rather than continuously selecting different actions to explore. This behaviour was found quite helpful in maintaining a more consistent policy over time for the online interactions in the museum, and avoiding that, at every interaction, the robot explores completely new actions. As it will be discussed in 5.5.3, in this real-world scenario, completely exploring all the states and actions is unfeasible. Therefore, exploring as many actions as possible at the beginning of the learning would not have allowed the policy to stabilise into a consistent behaviour for inter-

Algorithm 2: Online learning algorithm

```
Data: S, A, H, C_s ∀ s ∈ S
Q_h(s,a) ← 0 ∀ (s,a) ∈ S × A and h = 1,...,H;
N_h(s,a,s') ← 0 ∀ (s,a,s') ∈ S × A × S and h = 1,...,H;
D ← ∅;
foreach episode do
    // Policy optimisation
    for h = H,...,1 do
        for (s,a) ∈ S × A do
            Q̂_h(s,a) ← r(s) + E_{s'~p̂_h(·|s,a)}[V_{h+1}(s')];
            Q_h(s,a) ← Q̂_h(s,a) + bonus(h,s,a);
            V_h(s) ← min{H − h, max_{a∈A} Q_h(s,a)};
        end
    end
    π_h(s) ← arg max_{a∈C_s} Q_h(s,a) ∀ s ∈ S;
    // Policy execution
    Observe s_1;
    for h = 1,...,H do
        Execute a_h ← π_h(s_h);
        Observe r_h and s_{h+1};
        Add (s_h, a_h, s_{h+1}) to D
    end
    // Model update
    N_h(s,a,s') ←| {(s_h, a_h, s_{h+1}) ∈ D : (s_h, a_h, s_{h+1}) = (s,a,s')} | ∀ (s,a,s',h);
    p̂_h(s' | s,a) ← N_h(s,a,s') / |N_h(s,a,·)| ∀ (s,a,s',h);
end
```

acting. However, in different scenarios where it is easier to perform exploration, this effect could be beneficial as it could lead to finding the best policy quicker. In that case the studied effect can be eliminated by *optimistically* initialising $Q_h(s,a) = H - h$ $\forall\,(s,a) \in S \times A$ and $h = 1, \dots, H$, assuming that at the beginning all the unexplored state-action pairs will give the best possible return.

Algorithm 3: `bonus`

```
Data: N_h(s,a,s'), H
N ←| N_h(s,a,·) |;
b ← sqrt(1/N + (H−h)/(σ·N));
return min{b, H − h}
```

5.5 Experiments

In order to validate the hypothesis that optimising a social robot policy to maximise the user engagement leads to more sustained human-robot interactions, a long-term study in the museum with our tour guide robot Lindsey was performed. Since 2019, Lindsey has been delivering the guided tours to the visitors in a "static" way, as described in Chapter 3. With the static policy, in each tour the robot would guide the users to the exhibits always in the same order, and users were allowed to choose to receive a more detailed description of each item. Here instead, the users' engagement value is employed as a guiding metric for selecting actions during the interactions. This choice is motivated by the fact that users' displaying different perceived engagement have different willingness to continue the interaction [71], as also confirmed by our data in Figure 5.8(a) where for example, a perceived low engagement at the first stop of the tour means a 20% decrease in chances of users continuing the interaction. In particular, the policy is allowed to 1) choose freely the order of items visited in the tours; and 2) decide whether to provide additional information about each specific item rather than asking the user. The model of our scenario, namely the visitation frequency table $N_h(s, a, s')$, is initialised with the transitions from the thousands of episodes collected with the static policy. During the learning phase, the *optimistic* learning algorithm implemented takes care of directing the exploration toward the other less explored areas. The reward function is computed online from the engagement values detected by our engagement model, as described in Subsection 5.4.2.

The robot deployment was interrupted in 2020 because of the COVID lockdown and resumed in December 2021, the date when the learning phase started. During the following eight weeks the robot behaviour was driven by the online learning algorithm, which explored the different actions it could now take while exploiting the model to increase users' willingness to continue. In order to assess more appropriately the effect of the learned policy on the tour success and on users' engagement, a 2-week verification phase is performed afterwards, during which the original static policy again drives the robot behaviour. This validation was necessary in order to observe whether the adaptation algorithm actually caused the effect over our results during the learning phase or it was due to other spurious effects, like people willing to spend more time at the museum after COVID lockdowns or face mask wearing.

5.5.1 Performances of Learned Policy

This section reports the results of our evaluation to assess the performances of the learned robot behaviour. Figure 5.5 shows the evolution of different metrics over the entire deployment (omitting the periods during which the robot was not operational). The results are grouped and averaged per-week, with the number of tours for each data point reported in Figure 5.5(a) for significance. During the deployment period, the museum scenario did not change in any significant way to affect the robot tour

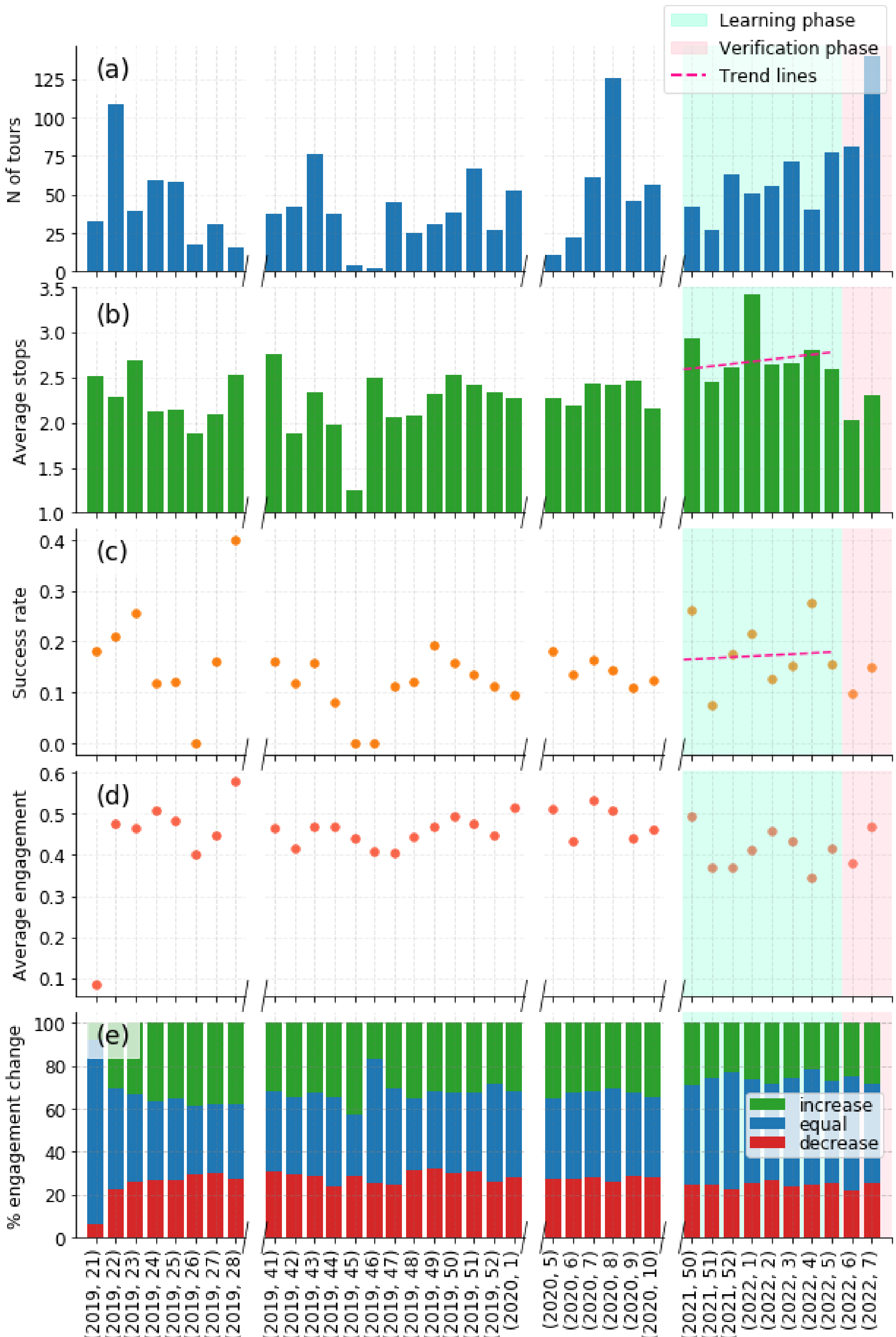


Figure 5.5: (a) Number of guided tours delivered per week, (b) average number of stops visited in the tours, (c) rate of tours completed to the end, (d) average scalar engagement level during tours (e) average change in engagement binned value (i.e. from `HIGH`, `MEDIUM`, `LOW`) between each pair of consecutive actions. Missing time ranges are caused by the robot not being operational or museum closures.

guides. Similarly, the robot structure and touch screen interface had only minor interventions like the replacement of an RGB-D sensor. The data reported for the weeks from (2019, 21) to (2020, 10) and after (2022, 5) are results of the "static" tours, while from (2021, 50) to (2022, 5) are caused by the learned policy. Figure 5.5(b) reports the average number of stops visited (exhibits described) in the tours. During the learning phase, an increase of 22.8% over the previous period is reported, with the verification phase bringing back the values in line with the previous data for the static tour. Similar results are obtained for the tour success rate, i.e. the rate of tours that terminate after all the stops have been visited, with an overall increase of over 30% over the static tour, as shown in Figure 5.5(c). Notice that the nominal number of items is 6 for the `Death` tour and 5 for the rest; therefore, a lower number of items means that the users have stopped or abandoned the interaction before the end. Trend lines, computed as the linear fit to the data, are shown for the average number of stops and the success rate of the learning period. The p values of the F-statistics versus a constant model are $2.7\mathtt{e}{-}6$ and 0.03 respectively for the two metrics, which indicates that the increasing numbers reported over time as the learning continues are statistically significant with $p < 0.05$. These results, taken together, suggest that our learned policy can indeed lead users to keep a more sustained engagement during the interactions in the museum, confirming the hypotheses which motivate this work.

In addition, the users' engagement detected by the regression model is studied here in order to understand whether the changes in robot behaviour directly influence it. The plots in Figure 5.5(d)-(e) report the average engagement and the average change in engagement over consecutive actions during the guided tours. This data suggests that during the learning phase, there was a general decrease in detected engagement over the previous and the verification periods. Morevoer, the learned policy maintained the users' engagement stable over time almost 50% of the time, unlike the previous period where there was a nearly equal chance of increasing, decreasing or maintaining the engagement stable.

5.5.2 Analysis of Learned Tours

While the previous section reported the effects of the behavioural adaptation on the guided tour performances and the users' engagement, this section analyses how the guided tour itself has changed after the learning phase compared to the initial static policy. One of the effects of allowing the learning algorithm to explore different actions is that now the robot can scramble the sequence of the items in the tour, guiding people to places in different orders. Figure 5.6 shows how the final learned policy would conduct the `art` tour at different levels of engagement detected during execution, as a representative example of the adaptation. The figure evidences how each different users' engagement levels generates a different tour. Additionally, the different engagements had the effect of varying the amount of information given to the users at each exhibit. Figure 5.7 reports how many times an additional description

Figure 5.6: Example of tour learned after behavioural adaptation. This figure shows the different actions that the learned policy would for the tour `art` if we were to fix the engagement level at `LOW`,`MEDIUM` and `HIGH` throughout the interaction. The blocks "Item N" represent the actions of guiding visitors to item N and giving a brief description of it; the blocks labeled with "Details" represent the action of describing the item in the previous block with more details.

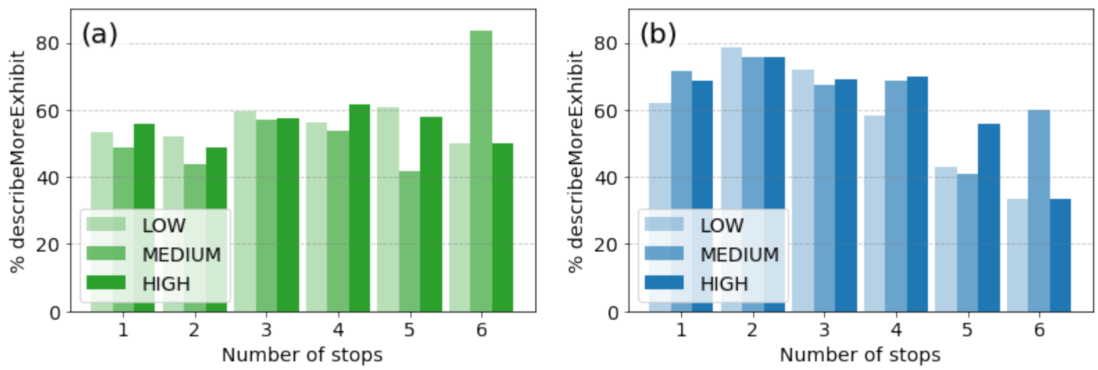


Figure 5.7: Percentage of times additional information is provided to the user at different steps of the tour and for different engagement values. (a) For static tours, users request the additional information; (b) the learned policy decides when to give more information.

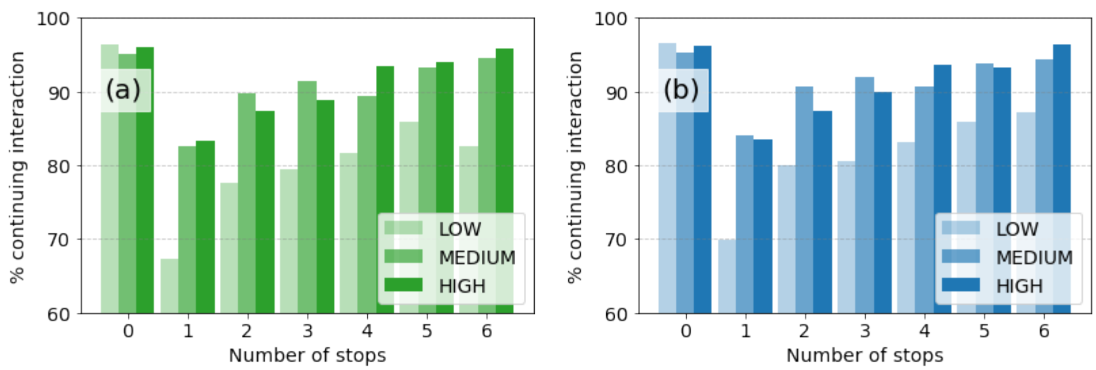


Figure 5.8: Percentage of times the tour is continued, rather than stopped or abandoned, at the different number of items visited in the tour and for different engagement values. (a) Static tour, (b) learned tour. Stop 0 corresponds to the `describeTour` action before reaching any exhibit.

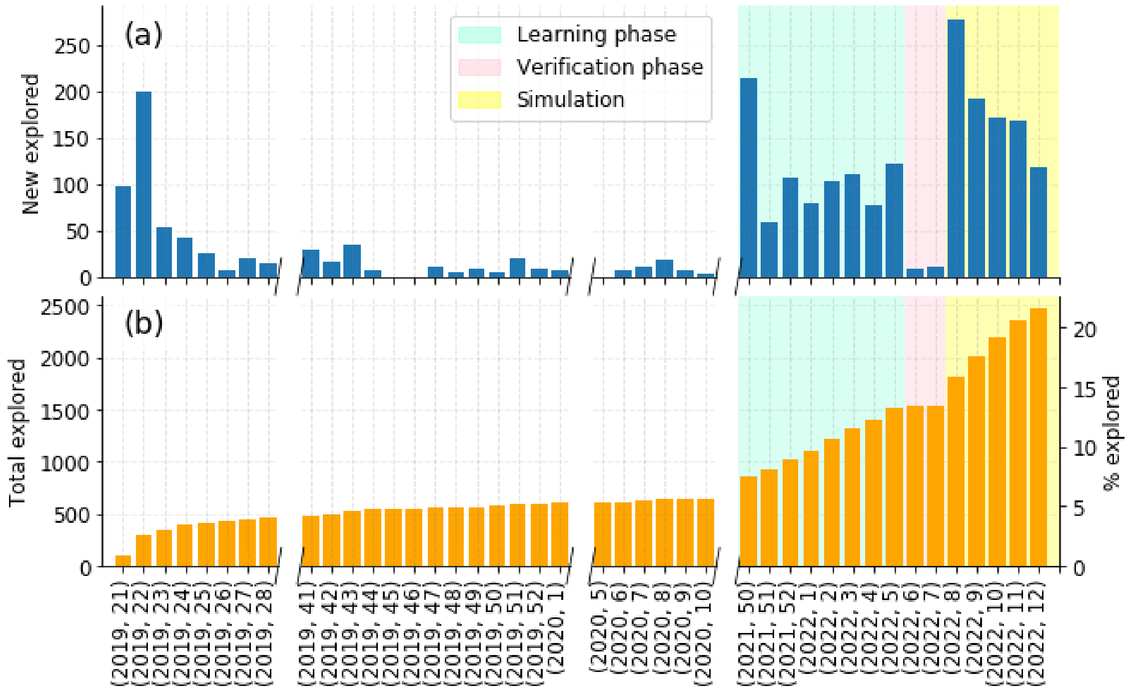


Figure 5.9: Exploration of the state-action space of the policy per week. (a) Number of newly explored state-action pairs, (b) cumulative exploration.

was given to the users by performing the `describeMoreExhibit` action for each stop number of the tour at different engagement levels. It can be observed that the robot has learned to give more details more often at the beginning of the tour, while for later steps only when the engagement is either at a `MEDIUM` or `HIGH` level.

Finally, an analysis is performed to verify whether different engagement levels correlate with different willingness to remain in the interaction. Figure 5.8 shows the percentage of times users have continued the interaction at different levels of engagement before and after the robot behaviour adaptation. From the figure, it can be observed that users are more prone to disengage in earlier stops, after the initial tour description, and when they have an engagement level which is `LOW`, consistent with the intuition that poorly engaged people are less willing to continue the interaction in the first place. After the learning phase, there is an overall increase in the probability of continuing the interaction compared to the static policy condition, but no other significant difference can be observed.

5.5.3 State-Action Exploration

The UCBVI algorithm for adapting the robot behaviour gives a simple yet efficient way of exploring new unseen states while exploiting the best performing actions, as described in Subsection 5.4.3. This section hence aims to analyse how the algorithm has effectively explored the state-action space in our scenario.

Figure 5.9 shows the amount of exploration the robot has done, reported per week. Following our expectations, the static tour performs little to no exploration, given that the robot always chooses the same actions at the same point in the tour. Note that in this condition, a small amount of exploration is still happening. However, it is entirely governed by the users' behaviour during the tour, i.e. asking for more information or not and manifesting different levels of engagement. Differently, the adoption of our learning algorithm has brought an initial week of high exploration of the many state-action transitions that were previously forbidden and kept a steady exploration rate in the subsequent weeks. The exploration is reduced in the weeks following the first, given that the policy at that point started to exploit the new actions that have resulted in being more promising in terms of future users' engagement, as also evidenced by the analysis of the learning outcomes in Subsection 5.5.2 and Subsection 5.5.1.

The exploration rate achieved as of the beginning of 2022 is only about 13% of the entire state-action space. Therefore, it is expected that the current robot policy can be further improved by continuing to explore the environment when the learning phase is resumed after the verification period. However, it can be hypothesised that the number of states that the robot can explore will be less and less as it continues to act in the environment, making it impossible for it to perform a complete exploration. As was mentioned in the previous paragraph, some parts of the state-action space are not directly explorable by the robot but can only be observed indirectly as a consequence of the user behaviour. In order to validate this hypothesis, a simulation of the robot acting in the environment was performed for the successive five weeks into the future, during which the learned policy is used to act while being improved over time, and the observed states are provided by sampling uniformly from the set of all possible transitions at each point in the tour. The simulated exploration, shown in Figure 5.9, is an upper-bound of the empirical exploration that can be obtained in real life because it does not simulate the possibility of *stopping* or *abandoning* the interaction, which would ultimately prevent exploring further into the tour sequence. The data shows that as time passes the number of states that the algorithm is able to explore decreases, as expected from our intuition.

5.6 Discussion

The results of the proposed behavioural adaptation framework for Lindsey the tour guide shows that the continuous engagement assessment value from Chapter 4, is a valuable metric to drive learning. This observation follows previous work from Meng et al. [65] which used continuous engagement as a reward for an adaptive interactive museum exhibition. This validates **Hypothesis 3**, since it demonstrated that different levels of users engagement can provide a discriminating factor for understanding how much people want to be in the interaction (and therefore can be used to learn contingency policies). It has been observed that during the learning period both the duration of tours (in terms of number of items visited) and the rate of complete tours delivered increased over time. Given that a complete exploration of the state-action space was not achieved during the 2-months learning period (and it is possibly unfeasible anyway, as discussed in Subsection 5.5.3), this leads to the assumption that with long training time the performances of the robot would increase more and more. This shows the advantage of adopting the UCBVI learning algorithm which allows to progressively learn better policies without facing the problem of having to retrain the entire model on the new data (which poses limitations in terms of training time and storage) of more data-hungry approaches, such as Deep-RL methods [74]. Moreover, the devised approach allowed the definition of constraints in the sequence of actions the robot could execute, and learn, which ensured that at any point during the learning period the robot behaviour would remain consistent. These results fulfilled **Objective 4**, defined in Section 1.1, and satisfied **Hypotheis 4** under the assumption that the users deciding to spend more time in the interaction with the robot is a manifestation of an improved quality of the interaction offered by the robot behaviours.

This chapter presented the results of a long-term deployment that spanned over a period of almost three years, during which a COVID pandemic happened which forced the museum to stay closed for about one year and seven months, therefore interrupting the robot operations during that period. Over such an extended period, it is reasonable to think that people's behaviour has changed, in particular after the COVID lockdowns and precautions which became part of everyone's everyday life, with the most apparent change being wearing face masks in the museum. For this reason, and because the learning phase started just after the museum re-opened, the verification phase was necessary in order to rule out spurious effects that could be biasing the improved efficacy of the robot's guided tours. Possibly, a visible effect of the changes in users' behaviour is visible in Figure 5.5(d-e) where an overall decrease in the users' detected engagement level can be observed. This could be caused by the effect that face cover wearing has on the engagement regression model. Such a hypothesis could explain the reason why the robot performances in sustaining interactions during the guided tours have improved, even though the users' engagement seems to have decreased during the same period.

5.7 Summary

This chapter presented an approach for adapting the robot behaviour during real-world social interactions in the context of robot-guided tours in a public museum. The learning algorithm is based on a Reinforcement Learning (RL) method, the Upper Confidence Bound Value Iteration (UCBVI), which efficiently balances exploration with exploitation allowing the robot to integrate past experience and to explore new states and actions in an informed way in the deployment. The method is also able to learn online, updating the policy at the end of every tour guide.

The optimisation of the robot policy is performed by maximisation of users' engagement displayed during the interactions. Such users' engagement is assessed by the regression model described in Chapter 4. This metric gives a way to holistically assess the quality of the interaction without resorting to asking for direct feedback to the user or performing manual annotation, therefore enabling online learning during the interactions.

With the experimental validation performed, it can be observed that the proposed framework leads to more extended guided tours with the robot, and the users stop or abandon the interactions less frequently than before adaptation. Such results are achieved after only a couple of months of learning from real interactions in the museum, when only a fraction of the entire state-space has been explored. Therefore, it is expected that with the continuation of the ongoing deployment, the robot policy will get better and better over time.

CHAPTER 6

Conclusions

THIS final chapter provides a highlight of the outcomes of the work that has been described in this thesis, drawing conclusions into how such achievements are relevant to the objective of the research project and to the wider HRI community. In particular, the next section will take the *Objectives* and *Hypotheses* defined in Chapter 1 as a guiding template to analyse how the solutions proposed throughout the rest of the chapters have addressed them, focusing on evidencing what has worked well and whether the hypotheses have been validated. Finally, the issues that are yet to be solved and the new challenges that lie ahead for future social robotics applications are described by briefly outlining possible research directions to address them.

6.1 Project Outcomes

6.1.1 A Robust and Autonomous Social Robot

The first tangible outcome of the work of this thesis is enabling the deployment of an autonomous social robot in a public space, The Collection museum[1], for long periods of time. The empirical results shown in Chapter 3 provide evidence of this claim by reporting that Lindsey the robot was deployed for a total duration of 3.3 years, with interruptions caused by hardware malfunctions and COVID-19 restrictions bringing down the total amount of the robot being actually operational to 1.2 years, and has travelled for more than 11 hundreds of kilometres. During this extended period of autonomous operations, the robot was effectively interacting with users for more than 75% of its time, showing that the system was functional and available for the museum visitors. To the best of the knowledge of the author of this thesis, these results show that the system is among the best performing in the HRI literature concerning autonomous social robots deployed in public spaces. Moreover, it is worth highlighting that the deployment, which started in 2018, is still ongoing to date, and there are currently no plans to end it in the foreseeable future. Moreover,

[1] https://www.thecollectionmuseum.com/robot-at-the-collection

the museum's operator found the robot's presence in the gallery an added value to the cultural experience they provide to the public, rather than representing a nuisance for them to manage or a distraction for the users to enjoy the visits. This achievement was made possible by the efforts to maximise the self-sufficiency of the robot during its operations, taking into account along the way the feedback from the museum's staff and the users. Section 3.5 describes the approaches used, which include various recovery strategies, interfaces for monitoring and managing the robot during its operation and a system for the notification of critical events which require immediate intervention. All these efforts allowed the museum employees and us to only pay attention to the robot when real manual help was necessary while offloading most of the other autonomy issues to the robot system itself or the users. This, overall, enabled the robot to perform autonomously in the long-term deployment. Based on this discussion, it can be established that **Objective 1** – i.e. *"Deployment of an autonomous robot to provide guided tours to visitors in a public museum. The robot framework must be robust to the point that it enables Long-Term Autonomy (LTA) requiring minimum in-situ assistance from experts."* – was satisfied.

Having a robust robotic system deployed in a social scenario enables the study of various phenomena that happen in real-world environments, in particular with **Hypothesis 1** this thesis verifies the claim *"Robot dynamic behaviours are more effective than static behaviours at maintaining users' engagement throughout the interactions"*. In order to study this effect, a long-term two-stage study has been implemented where first in Section 3.6 the usage patterns of the robot, which performs a static hand-crafted guided tour behaviour, by the museum visitors is studied in order to ascertain whether it's true that people do not find static robot behaviour very engaging. This analysis satisfied the **Objective 2**: *"Analysis of the initial deployment –with a static tour– specifically focused at observing the visitors' engagement with the technology during the guided tours."* and showed that it is indeed true that people typically quickly disengage from the interactions with Lindsey in this condition. Secondly, in Section 5.5, the users' engagement is reported when the visitors interact with the same robot behaviour, which now learns over time to take into account the users' state during the interaction in order to adapt their behaviour. The first and the second conditions are compared, with the results reported on a weekly basis given that the adapting behaviour learning continues to improve over time, showing that the users prefer to continue interacting with the robot with the adapting behaviour more often than with the static tour. More discussion about the efficacy of the adapting behaviour can be found in the following sections. Here, it is essential to highlight that the long-term analysis of an adaptive system that requires months of real-world interactions in a public space to learn was enabled by the autonomy and robustness of the entire system, as discussed above.

6.1.2 A Ready-to-Use Model of Users Engagement

The discussion in Chapter 2, which gives a general description of the adaptation framework implemented to improve the social interactions between robots and their users, highlights the importance for the robot to be able to estimate the users' engagement from its point of view and in real-time during the interactions. This requirement, which stems from the need to obtain a quick and inexpensive assessment of the robot's behaviour, led to establishing a ground truth of the users' engagement from the robot camera that can be easily translated into a guiding signal for learning. With the collection of the interactions data, the human annotation procedure for the TOGURO dataset, and the evaluation of the inter-rater agreements, as described in Section 4.4, it was possible to validate **Hypothesis 2** which states that *"Humans can intuitively and holistically assess the engagement of the people interacting with them, being able to give a continuous engagement score from first-person view observations"*. The evaluation found an overall medium to high agreement between the independent coders, highlighting that they were, in fact, annotating the same quantity even though they did not receive a specific definition of engagement to comply with but were free to use their own intuitive judgement. The approach used here, of annotating a concept that is difficult to formalise using the people intuitive assessment, is novel in the specific application to the measurement of engagement (it was previously applied to assessing the "quality of interaction" in Tanaka et al.[91]) and with the results of this work it was proven to be effective. This approach is in contrast with most previous works which try to identify specific features of engagement (like cues or behaviours) assuming that they can directly explain the phenomenon being assessed. The results from this work open the door to the use of such a continuous audience response method for annotating abstract phenomena like engagement in future social robotics work, given that it allows to annotate a large number of videos without spending more time than what is needed to watch each video once and, at the same time, reduces training time for the coders.

Once a dataset with a reliable ground truth has been created, Subsection 5.4.2 describes a deep regression model trained from such data that can be deployed on a robot with modest computing requirements for assessing, in real-time, the users' engagement during the interactions. This provides a solution to the **Objective 3** – i.e. *"devise a model of users engagement with the technology, trained from human engagement annotations, that can be used as a proxy for automatically evaluating the robot behaviour as a measure of the quality of the interactions it delivers to the users."*. In a social context, like a robotised guided tour in the museum, the principal objective of the robot is to engage with its audience so that it can effectively entertain them and educate them about the museum's exhibition. For this reason, in this thesis, the users' engagement assessment is considered as the metric to estimate the quality of the interaction and, in turn, seen as feedback for the robot's behaviour. The experimental results testing the model's prediction against the ground truths from the TOGURO dataset and the UE-HRI data [12] show that such model trained only from the holistic assessment of human coders accurately predicts the users' engagement in both the museum scenario and in different environments with different robots and cameras. Given the broad applicability of the method, requiring only a camera on the robot board, little computational abilities and no re-training or tuning of the model, the trained engagement model has been openly released[2] for the benefit of other social robotics applications.

6.1.3 An Online Adaptation Framework for Social Robots

The final contribution of this thesis, which is the focus of Chapter 5, is a learning component which, putting together the robotic system with the guided tours behaviour described in Chapter 3 and the continuous engagement assessment model presented in Chapter 4, is able to adapt the robot behaviour with the goal of maximising the users' engagement during interactions. The novelties of the learning approach proposed are 1) the use of the holistic users' engagement assessment as a metric to drive the learning, and 2) the application of an optimistic RL approach to adapt social robot behaviours enabling to learn efficiently from the interactions in the real world. Such a learning framework provides a solution for the **Objective 4** – i.e. *"Implementation of a RL algorithm that optimises the "robotised" guided tours policy by maximising the users' engagement over the entire duration of the interaction"*. This framework was implemented and deployed on the Lindsey mobile robot and tested in the real scenario of the museum; however, the approach's only assumptions are that it can detect engagement from the robot's camera and that the interactions are episodic. Therefore, it is applicable in most social robotics applications where the robot is equipped with a camera facing the users during the interactions, i.e. enabling to detect engagement using the regression model of Chapter 4, and each interaction is self-contained with a beginning and end. Additionally, the devised RL algorithm gives a natural way for incorporating an initial prior of the environment dynamics which, in this case, consists of the exploration performed over the years in the museum with a static behaviour. However, such a prior did not prevent the robot from efficiently exploring different actions and new states.

The results of deploying the proposed adaptation algorithm, shown in Section 5.5, validates the approach reporting that as the policy improved over time, the users spent more time in the interaction, i.e. stopping or abandoning the interaction later in the tour on average. Moreover, they were more willing to perform a complete guided tour, i.e. decided to stop or abandon the tour less frequently in the first place. These results, validated by a statistical analysis over eight weeks of adaptation and hundreds of tours with real visitors of the museum, lead to confirming the ideas that *"The engagement level displayed by users indicates their willingness to stay in the interaction"* and that *"by taking into account such users engagement, a robot can improve the quality of the interactions"*–i.e. **Hypotheses 3 and 4**.

[2] https://github.com/LCAS/engagement_detector

6.2 Limitations and Future Work

This section outlines the limitations of the approaches and methodologies proposed in this thesis for developing a tour guide robot to interact autonomously with museums' visitors in a socially appropriate way. In the following, each distinct contribution is analysed separately and possible future directions for work are proposed therein.

6.2.1 Long-Term Autonomy

Despite the long lasting deployment of Lindsey in the museum, totaling more than 1100 km of autonomous navigation and having performed more than 31000 tasks, a substantial amount of technical support was required throughout this period. Firstly, hardware failures are still common in robotics systems due to the co-existence of multiple specialist sensors and actuators. Integrating these components in a functioning system, at the beginning of the deployment or when components need to be replaced at a later iteration, requires substantial efforts which could be reduced by the use of standardised platforms and hardware setups available in the market. Hardware failures are also difficult to spot quickly until it generates an evident malfunction of the robot's operations, at which point it is usually too late; for example, a head camera malfunction during operations could remain unspotted until at the end of the day it prevents the robot from returning to the docking station to recharge. The Sentor node, described in Subsection 3.5.3, has been developed in this project with the purpose of rapidly discovering such malfunctions by allowing to define anomalous conditions which could represent failures. However, this system has two main limitations: 1) to define a monitoring condition, one must know beforehand about the problem and how it is manifested via ROS topics; and 2) the conditions we can define to match a failure, sometimes also match non-failure cases. Therefore, despite being able to automate the identification of known failures, it cannot automatically identify newly arisen anomalies and it requires the roboticists to analyse the system's notifications to ascertain whether they represent genuine failure cases. In future works, automatic detection of failures could be a promising direction to explore for solving these limitations. Recent works using DL methodologies have shown to be able to detect anomalies from time series data [4, 97].

Autonomous navigation during the presented deployment was limited by the need of keeping the robot's metric map continuously up-to-date due to changes in the museum's temporary exhibitions. The localisation algorithm deployed on Lindsey, the standard ROS amcl algorithm, assumes a static and accurate map of the environment is available to match the sensor's observations and find the robot's location. However, delays in communicating the museum's changes and the time required to scan anew the environment after meant that the robot was being left stopped at the station for a few days at times and that, at other times, it was navigating with an out-of-date map (hence, probably causing more navigation errors).

Finally, robot deployments in public spaces would greatly benefit from more intuitive human-robot/human-computer interfaces allowing non-expert users to observe the robot's state during operations and modify, when needed, its actions. This would benefit the explainability of the robot by giving end users a more intuitive understanding of the robot's actions and the museum managers the possibility of personalising the robot experience to reflex the day-by-day changes in the museum's exhibitions. This is particularly important for managing failures and unexpected robot behaviours, in order to avoid the user loss of trust in the robot technology and to be able to explain the failure, with possible recovery strategies, in a way that is understandable by non-experts. Explainable AI is a promising direction that should be explored in the future for explaining failures to the general user [24].

6.2.2 User Engagement Assessment

In Chapter Chapter 4, it has been shown how assessing the users' engagement in real-time from the robot's point of view is possible with an end-to-end trained model. This model, trained on data from the general population makes assessments holistically; i.e., not focussing specifically on particular features or demographics, and assessing the entire interacting group of users. While this was a design feature of the described work, enabling continuous and reliable assessments during operations, it has its own drawbacks which should be addressed in future works. In the first instance, we know from the results shown in Figure 4.7 that the model performances degrade for specific user demographics. While this could be a result of a bias in the training data, it is also possible that engagement estimation for specific user groups is more challenging than for others. It is not clear also whether choosing different timescales for integrating the input data benefits the estimation for particular user groups. Moreover, it is possible that certain demographics were not present at all in the training data, and therefore, the model is not able to estimate their engagement accurately. For example, it is known that people with intellectual disabilities and autistic people engage in interactions differently than the general population, with some displaying a limited ability to show engagement and emotions in general [85].

In future work, more research is needed to understand how the quality of engagement assessment changes for different demographics and to what extent a model trained on the general population can be used in specific cases. With the lack of a commonly agreed definition for engagement [36], empirical evaluations are needed. If a general model of engagement for different demographics is not achievable, because of individual differences in expressivity, it is possible that personalisation is the direction to aim toward, such as following from the work of Won Park et al. [98] which learns personalised engagement estimation models from human expertise. A different option is the generation of persona-based profiles that can tailor the engagement assessment based on different demographic groups, taking inspiration from Andriella et al. [2]. As highlighted by Salam and Chetouani [81], context is an important factor in engagement. To date however, besides evaluations on specific settings and datasets, there is still a lack of generalised computational models that can integrate contextual information in real-time for the assessment of engagement in real-world situations.

6.2.3 Behavioural Adaptation

An important aspect of robotics behavioural adaptation is the ability to change the actions the robot performs in accordance with the users' preferences and state during the interactions. In the approach proposed in Chapter 5, while the robot learned to improve its decision-making based on the immediate users' engagement level, it did not address the problem of taking into account their individual preferences or their physical and social state beyond engagement. Due to the application scenario of the approach – a tour guide robot in a public museum – most interactions involved users who encountered the robot for the first time, with little to no repeated interactions with the same users. This limited the possibility of personalising the robot's behaviour to specific users, with learning possible only when considering the visitors' population as a whole and trying to generalise for the entire group or to specific demographics (again possibly generating persona-based profiles [2]). There are many interleaving causes that may lead people to disengage with a robot, some of which are not directly attributable to the robot's behaviour. For example, people may need to leave an interaction because they have a more urgent matter to attend, because they feel unwell that day, or because of a negative pre-conceived attitude towards robots. From the results on the quality of the interactions during the learning phase of the proposed approach, we can observe that while the success rate and the average number of stops attended during the tours increased, the average engagement value of users seemed to reduce over time. Additional studies are required to appropriately study this effect and discern its causes.

A limitation of this work is that the learning model is not able to address the situation in which the behaviours or preferences of visitors change over time. The UCBVI bonus decreases over time as a certain state and action pair continues being explored, while assuming that the value estimation gets closer and closer to the real one. However, in non-stationary environments, i.e. when visitors change over time, this assumption is not valid anymore. A general idea to take into consideration this aspect in future research is to modify the bonus function so that it can also increase in situations where the divergence between the current estimate and the newly encountered values is high.

A second factor limiting the applicability of the proposed framework for more complex problems is that it cannot handle high-dimensional state and action spaces. Recent efforts, such as Ciosek et al. [23], have provided model-free methods for optimistic Deep RL approaches which are able to improve sample efficiency in continuous control tasks, however, such methods still require a prohibitive amount of exploration (in the order of millions of steps) for simulated environment problems, making them unfeasible for real-world human-robot interactions problems. Therefore, more research is needed to study how function approximation methods could be used to handle high-dimensional data in RL problems like the one presented here.

References

[1] A. Andriella, C. Torras, C. Abdelnour and G. Alenyà, 'Introducing CARESSER: A framework for in situ learning robot social assistance from expert knowledge and demonstrations,' *User Modeling and User-Adapted Interaction*, pp. 1–56, 2022, ISSN: 0924-1868. DOI: 10.1007/s11257-021-09316-5. [Online]. Available: https://link.springer.com/article/10.1007/s11257-021-09316-5 (cit. on p. 77).

[2] A. Andriella, C. Torras and G. Alenya, 'Learning Robot Policies Using a High-Level Abstraction Persona-Behaviour Simulator,' in *2019 28th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2019*, 2019, ISBN: 9781728126227. DOI: 10.1109/RO-MAN46459.2019.8956357 (cit. on pp. 76, 104, 105).

[3] M. G. Azar, I. Osband and R. Munos, 'Minimax regret bounds for reinforcement learning,' in *ICML*, 2017 (cit. on pp. 81, 87).

[4] D. Azzalini, L. Bonali and F. Amigoni, 'A minimally supervised approach based on variational autoencoders for anomaly detection in autonomous robots,' *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2985–2992, 2021. DOI: 10.1109/LRA.2021.3062597 (cit. on p. 103).

[5] S. S. Balasuriya, L. Sitbon, M. Brereton and S. Koplick, 'How can social robots spark collaboration and engagement among people with intellectual disability?' In *ACM International Conference Proceeding Series*, 2019, pp. 209–220, ISBN: 9781450376969. DOI: 10.1145/3369457.3370915. [Online]. Available: https://doi.org/10.1145/3369457.3370915 (cit. on pp. 48, 76).

[6] T. Baltrušaitis, M. Mahmoud and P. Robinson, 'Cross-dataset learning and person-specific normalisation for automatic Action Unit detection,' *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2015*, vol. 2015-January, 2015. DOI: 10.1109/FG.2015.7284869 (cit. on p. 51).

[7] A. G. Barto, 'Intrinsic motivation and reinforcement learning,' in *Intrinsically Motivated Learning in Natural and Artificial Systems*, vol. 9783642323, Springer-Verlag Berlin Heidelberg, 2013, pp. 17–47, ISBN: 9783642323751.

DOI: 10.1007/978-3-642-32375-1_2. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-32375-1{_}2 (cit. on p. 78).

[8] T. Baur, G. Mehlmann, I. Damian, F. Lingenfelser, J. Wagner, B. Lugrin, E. André and P. Gebhard, 'Context-aware automated analysis and annotation of social human-agent interactions,' *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 5, no. 2, p. 11, 2015 (cit. on p. 57).

[9] P. Baxter, F. Del Duchetto and M. Hanheide, 'Engaging learners in dialogue interactivity development for mobile robots,' in *International Conference on Educational Robotics (EDUROBOTICS)*, 2018 (cit. on pp. 48, 75).

[10] P. Baxter, J. Kennedy, A.-L. Vollmer, J. de Greeff and T. Belpaeme, 'Tracking gaze over time in HRI as a proxy for engagement and attribution of social agency,' in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14*, New York, New York, USA: ACM Press, 2014, pp. 126–127, ISBN: 9781450326582. DOI: 10.1145/2559636.2559829 (cit. on pp. 46, 48, 50).

[11] R. Bellman, 'A markovian decision process,' *Indiana Univ. Math. J.*, vol. 6, pp. 679–684, 4 1957, ISSN: 0022-2518 (cit. on p. 80).

[12] A. Ben-Youssef, C. Clavel, S. Essid, M. Bilac, M. Chamoux and A. Lim, 'UE-HRI: a new dataset for the study of user engagement in spontaneous human-robot interactions,' in *Proceedings of the 19th ACM International Conference on Multimodal Interaction - ICMI 2017*, New York, New York, USA: ACM Press, 2017, pp. 464–472, ISBN: 9781450355438. DOI: 10.1145/3136755.3136814 (cit. on pp. 46, 47, 64, 67, 84, 101).

[13] A. Ben Youssef, G. Varni, S. Essid and C. Clavel, 'On-the-fly Detection of User Engagement Decrease in Spontaneous Human-Robot Interaction,' *International Journal of Social Robotics*, Jan. 2019. [Online]. Available: https://hal.telecom-paris.fr/hal-02288044 (cit. on pp. 51, 70).

[14] K. Best, 'Making museum tours better: Understanding what a guided tour really is and what a tour guide really does,' *Museum Management and Curatorship*, vol. 27, no. 1, pp. 35–52, 2012 (cit. on p. 42).

[15] J. Biswas and M. Veloso, 'The 1,000-km challenge: Insights and quantitative and qualitative results,' *IEEE Intelligent Systems*, vol. 31, no. 3, pp. 86–96, 2016 (cit. on pp. 14, 75).

[16] O. Brown, L. Roberts-Elliott, F. Del Duchetto, M. Hanheide and P. Baxter, 'Abstract visual programming of social robots for novice users,' in *ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 154–156, ISBN: 9781450370578. DOI: 10.1145/3371382.3378271 (cit. on p. 25).

[17] S. Bubeck and N. Cesa-Bianchi, 'Regret analysis of stochastic and nonstochastic multi-armed bandit problems,' *Foundations and Trends in Machine Learning*,

vol. 5, no. 1, pp. 1–122, 2012, ISSN: 19358237. DOI: 10.1561/2200000024. arXiv: 1204.5721 (cit. on p. 81).

[18] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner and S. Thrun, 'The interactive museum tour-guide robot,' in *Aaai/iaai*, 1998, pp. 11–18 (cit. on p. 15).

[19] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva and P. W. McOwan, 'Detecting engagement in hri: An exploration of social and task-based context,' in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, 2012, pp. 421–428. DOI: 10.1109/SocialCom-PASSAT.2012.51 (cit. on p. 50).

[20] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva and P. W. Mcowan, 'Context-sensitive affect recognition for a robotic game companion,' *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 4, no. 2, pp. 1–25, 2014 (cit. on p. 50).

[21] G. Castellano, A. Pereira, I. Leite, A. Paiva and P. W. McOwan, 'Detecting user engagement with a robot companion using task and social interaction-based features,' in *Proceedings of the 2009 international conference on Multimodal interfaces*, 2009, pp. 119–126 (cit. on pp. 48, 50).

[22] G. Castellano and C. Peters, 'Socially perceptive robots: Challenges and concerns,' *Interaction Studies*, vol. 11, no. 2, pp. 201–207, 2010, ISSN: 1572-0373. DOI: 10.1075/IS.11.2.04CAS (cit. on pp. 8, 15).

[23] K. Ciosek, R. Loftin, Q. Vuong and K. Hofmann, 'Better exploration with optimistic actor-critic,' in *Advances in Neural Information Processing Systems*, vol. 32, 2019. DOI: 10.5555/3454287.3454447. arXiv: 1910.12807 (cit. on p. 105).

[24] D. Das, S. Banerjee and S. Chernova, 'Explainable ai for robot failures: Generating explanations that improve user assistance in fault recovery,' in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21, Boulder, CO, USA: Association for Computing Machinery, 2021, 351–360, ISBN: 9781450382892. DOI: 10.1145/3434073.3444657. [Online]. Available: https://doi.org/10.1145/3434073.3444657 (cit. on p. 104).

[25] E. Dağlarlı, S. F. Dağlarlı, G. Ö. Günel and H. Köse, 'Improving human-robot interaction based on joint attention,' *Applied Intelligence*, vol. 47, no. 1, pp. 62–82, 2017, ISSN: 15737497. DOI: 10.1007/S10489-016-0876-X/TABLES/5. [Online]. Available: https://link.springer.com/article/10.1007/s10489-016-0876-x (cit. on p. 9).

[26] F. Del Duchetto, P. Baxter and M. Hanheide, 'Lindsey the Tour Guide Robot - Usage Patterns in a Museum Long-Term Deployment,' in *International*

Conference on Robot & Human Interactive Communication (RO-MAN), New Delhi: IEEE, 2019 (cit. on pp. 74, 75).

[27] F. Del Duchetto, P. Baxter and M. Hanheide, 'Are You Still With Me? Continuous Engagement Assessment From a Robot's Point of View,' *Frontiers in Robotics and AI*, vol. 7, p. 116, 2020, ISSN: 22969144. DOI: 10.3389/frobt.2020.00116. arXiv: 2001.03515. [Online]. Available: www.frontiersin.org (cit. on p. 74).

[28] F. Del Duchetto and M. Hanheide, 'Learning on the Job: Long-Term Behavioural Adaptation in Human-Robot Interactions,' *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6934–6941, 2022. DOI: 10.1109/lra.2022.3178807. arXiv: 2203.10518 (cit. on p. 74).

[29] F. Del Duchetto, A. Kucukyilmaz, L. Iocchi and M. Hanheide, 'Do not make the same mistakes again and again: Learning local recovery policies for navigation from human demonstrations,' *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4084–4091, 2018 (cit. on p. 36).

[30] J. Duchi, E. Hazan and Y. Singer, 'Adaptive subgradient methods for online learning and stochastic optimization,' *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011 (cit. on p. 64).

[31] G. Durantin, S. Heath and J. Wiles, 'Social moments: A perspective on interaction for social robotics,' *Frontiers in Robotics and AI*, vol. 4, p. 24, 2017 (cit. on p. 50).

[32] A. El Korchi and Y. Ghanou, 'Unrestricted random sampling of data batch to improve the efficiency of neural networks,' in *Proceedings of the New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Society*, ACM, 2019, p. 26 (cit. on p. 64).

[33] M. E. Foster, A. Gaschler and M. Giuliani, 'How can I help you? Comparing engagement classification strategies for a robot bartender,' in *ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction*, 2013, pp. 255–261, ISBN: 9781450321297. DOI: 10.1145/2522848.2522879 (cit. on pp. 48, 51).

[34] ——, 'Automatically classifying user engagement for dynamic multi-party human-robot interaction,' *International Journal of Social Robotics*, vol. 9, no. 5, pp. 659–674, 2017 (cit. on p. 50).

[35] Y. Gao, F. Yang, M. Frisk, D. Hernandez, C. Peters and G. Castellano, 'Learning socially appropriate robot approaching behavior toward groups using deep reinforcement learning,' in *International Conference on Robot & Human Interactive Communication (RO-MAN)*, New Delhi: IEEE, 2019 (cit. on pp. 9, 77).

[36] N. Glas and C. Pelachaud, 'Definitions of Engagement in Human-Agent Interaction,' in *International Conference on Affective Computing and Intelli-*

gent Interaction (ACII), Xi'an, China: IEEE Press, 2015, pp. 944–949, ISBN: 9781479999538 (cit. on pp. 48, 104).

[37] B. Gleason and C. Greenhow, 'Hybrid education: The potential of teaching and learning with robot-mediated communication,' *Online Learning Journal*, vol. 21, no. 4, 2017 (cit. on pp. 41, 48).

[38] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz and J. Wang, 'Designing robots for long-term social interaction,' in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, IEEE Computer Society, 2005, pp. 1338–1343, ISBN: 0780389123. DOI: 10.1109/IROS.2005.1545303 (cit. on pp. 8, 15, 41).

[39] M. Hanheide, D. Hebesberger and T. Krajník, 'The when, where, and how: An adaptive robotic info-terminal for care home residents,' in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ACM, 2017, pp. 341–349 (cit. on pp. 14, 21, 75).

[40] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Mudrova, J. Young, J. Wyatt, D. Hebesberger, T. Kortner *et al.*, 'The strands project: Long-term autonomy in everyday environments,' *IEEE Robotics & Automation Magazine*, vol. 24, no. 3, pp. 146–156, 2017 (cit. on pp. 7, 14, 18, 26, 33, 41, 75).

[41] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory,' *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, ISSN: 08997667. DOI: 10.1162/neco.1997.9.8.1735. [Online]. Available: https://direct.mit.edu/neco/article/9/8/1735/6109/Long-Short-Term-Memory (cit. on p. 54).

[42] S. Hochreiter and J. Schmidhuber, 'Long short-term memory,' *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997 (cit. on p. 62).

[43] A. Holroyd, 'Generating engagement behaviors in human-robot interaction,' 2011 (cit. on pp. 50, 76).

[44] S. Honig and T. Oron-Gilad, 'Understanding and resolving failures in human-robot interaction: Literature review and model development,' *Frontiers in psychology*, vol. 9, p. 861, 2018 (cit. on p. 15).

[45] L. Iocchi, L. Jeanpierre, M. Lazaro and A.-I. Mouaddib, 'A practical framework for robust decision-theoretic planning and execution for service robots,' *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 26, no. 1, pp. 486–494, 2016. [Online]. Available: https://ojs.aaai.org/index.php/ICAPS/article/view/13790 (cit. on p. 20).

[46] T. Jaksch, R. Ortner and P. Auer, 'Near-optimal regret bounds for reinforcement learning,' in *J. Mach. Learn. Res.*, 2008 (cit. on p. 81).

[47] T. Kanda, T. Hirano, D. Eaton and H. Ishiguro, 'Interactive robots as social partners and peer tutors for children : A field trial,' *Human-Computer Interaction*, vol. 19, no. 1-2, pp. 61–84, 2004, ISSN: 07370024. DOI: 10.1207/s15327051hci1901&2_4. [Online]. Available: https://doi.org/10.1080/07370024.2004.9667340 (cit. on pp. 45, 75).

[48] D. Karreman, G. Ludden and V. Evers, 'Visiting cultural heritage with a tour guide robot: A user evaluation study in-the-wild,' in *ICSR 2015*, Springer LNAI, 2015, pp. 317–326 (cit. on p. 42).

[49] J. F. Kelley, 'An iterative design methodology for user-friendly natural language office information applications,' *ACM Transactions on Information Systems (TOIS)*, vol. 2, no. 1, pp. 26–41, 1984, ISSN: 15582868. DOI: 10.1145/357417.357420. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/357417.357420 (cit. on p. 13).

[50] A. Krizhevsky, I. Sutskever and G. E. Hinton, 'Imagenet classification with deep convolutional neural networks,' in *Advances in neural information processing systems*, 2012, pp. 1097–1105 (cit. on p. 62).

[51] A. Kubota, E. I. Peterson, V. Rajendren, H. Kress-Gazit and L. D. Riek, 'JESSIE: Synthesizing social robot behaviors for personalized neurorehabilitation and beyond,' in *ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA: ACM, 2020, pp. 121–129, ISBN: 9781450367462. DOI: 10.1145/3319502.3374836. [Online]. Available: https://doi.org/10.1145/3319502.3374836 (cit. on p. 76).

[52] L. Kunze, N. Hawes, T. Duckett, M. Hanheide and T. Krajník, 'Artificial intelligence for long-term robot autonomy: A survey,' *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4023–4030, 2018 (cit. on pp. 15, 75).

[53] S. Lathuilière, B. Massé, P. Mesejo and R. Horaud, 'Deep reinforcement learning for audio-visual gaze control,' in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 1555–1562 (cit. on p. 77).

[54] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, 'Backpropagation Applied to Handwritten Zip Code Recognition,' *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989, ISSN: 0899-7667. DOI: 10.1162/neco.1989.1.4.541 (cit. on p. 53).

[55] I. Leite, C. Martinho and A. Paiva, 'Social robots for long-term interaction: A survey,' *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013 (cit. on pp. 9, 15, 75, 77).

[56] I. Leite, C. Martinho, A. Pereira and A. Paiva, 'As time goes by: Long-term evaluation of social presence in robotic companions,' in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*,

2009, pp. 669–674, ISBN: 9781424450817. DOI: 10.1109/ROMAN.2009.5326256 (cit. on pp. 45, 75).

[57] I. Leite, M. McCoy, D. Ullman, N. Salomons and B. Scassellati, 'Comparing models of disengagement in individual and group interactions,' in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, 2015, pp. 99–105 (cit. on p. 50).

[58] S. Lemaignan, F. Garcia, A. Jacq and P. Dillenbourg, 'From real-time attention assessment to with-me-ness in human-robot interaction,' in *The eleventh acm/ieee international conference on human robot interaction*, IEEE Press, 2016, pp. 157–164 (cit. on p. 50).

[59] M. Leonetti and L. Iocchi, 'Learnpnp: A tool for learning agent behaviors,' in *Robot Soccer World Cup*, Springer, 2010, pp. 418–429 (cit. on p. 19).

[60] M. E. U. Ligthart, M. A. Neerincx and K. V. Hindriks, 'Memory-Based Personalization for Fostering a Long-Term Child-Robot Relationship,' *The ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, no. 2, pp. 80–89, 2022. [Online]. Available: https://bit.ly/3sdiQsd (cit. on p. 76).

[61] M. Luperto, J. Monroy, J. R. Ruiz-Sarmiento, F. A. Moreno, N. Basilico, J. Gonzalez-Jimenez and N. A. Borghese, 'Towards long-term deployment of a mobile robot for at-home ambient assisted living of the elderly,' in *2019 European Conference on Mobile Robots, ECMR 2019 - Proceedings*, 2019, ISBN: 9781728136059. DOI: 10.1109/ECMR.2019.8870924 (cit. on p. 14).

[62] M. L. McHugh, 'Interrater reliability: The kappa statistic,' *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012, ISSN: 13300962. DOI: 10.11613/bm.2012.031. [Online]. Available: /pmc/articles/PMC3900052//pmc/articles/PMC3900052/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/ (cit. on p. 59).

[63] E. Mcquillin, N. Churamani and H. Gunes, 'Learning Socially Appropriate Robo-waiter Behaviours through Real-time User Feedback,' *Proceedings of the 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, p. 2022, 2022. [Online]. Available: https://forms.gle/mE5FEBBJS133Y36KA (cit. on p. 77).

[64] W. Meeussen, E. Marder-Eppstein, K. Watts and B. P. Gerkey, 'Long term autonomy in office environments,' in *ALONE Workshop, In Proceedings of Robotics: Science and Systems (RSS'11), Los Angeles, USA*, 2011 (cit. on pp. 14, 75).

[65] L. Meng, D. Lin, A. Francey, R. Gorbet, P. Beesley and D. Kulić, 'Learning to Engage with Interactive Systems,' *ACM Transactions on Human-Robot Interaction*, vol. 10, no. 1, 2020, ISSN: 25739522. DOI: 10.1145/3408876 (cit. on pp. 77, 95).

[66] M. P. Michalowski, S. Sabanovic and R. Simmons, 'A spatial model of engagement for a social robot,' in *9th IEEE International Workshop on Advanced Motion Control, 2006.*, IEEE, 2006, pp. 762–767 (cit. on p. 50).

[67] M. Mori, 'The Uncanny Valley,' *Energy*, vol. 7, no. 4, pp. 33–35, 1970, ISSN: 10709932 (cit. on p. 8).

[68] O. Nocentini, L. Fiorini, G. Acerbi, A. Sorrentino, G. Mancioppi and F. Cavallo, 'A Survey of Behavioral Models for Social Robots,' *Robotics*, vol. 8, no. 3, p. 54, 2019, ISSN: 22186581. DOI: 10.3390/robotics8030054. [Online]. Available: https://www.mdpi.com/2218-6581/8/3/54/htmhttps://www.mdpi.com/2218-6581/8/3/54 (cit. on p. 9).

[69] I. R. Nourbakhsh, C. Kunz and T. Willeke, 'The mobot museum robot installations: A five year experiment,' in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, IEEE, vol. 4, 2003, pp. 3636–3641 (cit. on pp. 15, 42, 76).

[70] H. L. O'Brien and E. G. Toms, 'What is user engagement? a conceptual framework for defining user engagement with technology,' *Journal of the American society for Information Science and Technology*, vol. 59, no. 6, pp. 938–955, 2008 (cit. on p. 49).

[71] C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini and I. Poggi, 'A model of attention and interest using gaze behavior,' in *International Workshop on Intelligent Virtual Agents*, Springer, 2005, pp. 229–240 (cit. on pp. 49, 50, 89).

[72] C. C. Ponitz, S. E. Rimm-Kaufman, K. J. Grimm and T. W. Curby, 'Kindergarten classroom quality, behavioral engagement, and reading achievement,' *School Psychology Review*, vol. 38, no. 1, pp. 102–121, 2009 (cit. on p. 48).

[73] A. H. Qureshi, Y. Nakamura, Y. Yoshikawa and H. Ishiguro, 'Robot gains social intelligence through multimodal deep reinforcement learning,' in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, IEEE, 2016, pp. 745–751 (cit. on pp. 9, 70, 77).

[74] ——, 'Show, attend and interact: Perceivable human-robot social interaction through neural attention q-network,' in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2017, pp. 1639–1645 (cit. on pp. 52, 77, 96).

[75] C. Rich, B. Ponsler, A. Holroyd and C. L. Sidner, 'Recognizing engagement in human-robot interaction,' in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, 2010, pp. 375–382 (cit. on pp. 48, 50).

[76] S. Rosenthal, J. Biswas and M. Veloso, '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, 2010, pp. 915–922, ISBN: 9781617387715. DOI: 10.5555/1838206. [Online]. Available: www.ifaamas.org (cit. on p. 8).

[77] S. Rosenthal, M. Veloso and A. K. Dey, 'Is someone in this office available to help me?' *Journal of Intelligent & Robotic Systems*, vol. 66, no. 1-2, pp. 205–221, 2012 (cit. on p. 14).

[78] O. Rudovic, J. Lee, L. Mascarell-Maricic, B. W. Schuller and R. W. Picard, 'Measuring Engagement in Robot-Assisted Autism Therapy: A Cross-Cultural Study,' *Frontiers in Robotics and AI*, vol. 4, 2017, ISSN: 2296-9144. DOI: 10.3389/frobt.2017.00036 (cit. on p. 46).

[79] O. Rudovic, Y. Utsumi, J. Lee, J. Hernandez, E. C. Ferrer, B. Schuller and R. W. Picard, 'Culturenet: A deep learning approach for engagement intensity estimation from face images of children with autism,' in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 339–346 (cit. on pp. 51, 52, 62, 70).

[80] D. E. Rumelhart, G. E. Hinton and R. J. Williams, 'Learning representations by back-propagating errors,' *Nature*, vol. 323, no. 6088, pp. 533–536, 1986, ISSN: 00280836. DOI: 10.1038/323533a0. [Online]. Available: https://www.nature.com/articles/323533a0 (cit. on p. 53).

[81] H. Salam and M. Chetouani, 'A multi-level context-based modeling of engagement in human-robot interaction,' in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, IEEE, vol. 3, 2015, pp. 1–6 (cit. on pp. 49, 50, 104).

[82] ——, 'Engagement detection based on mutli-party cues for human robot interaction,' in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE, 2015, pp. 341–347 (cit. on p. 50).

[83] E. Senft, P. Baxter and T. Belpaeme, 'Human-guided learning of social action selection for robot-assisted therapy,' in *Proceedings of The 4th Workshop on Machine Learning for Interactive Systems at ICML 2015*, H. Cuayáhuitl, N. Dethlefs, L. Frommberger, M. Van Otterlo and O. Pietquin, Eds., ser. Proceedings of Machine Learning Research, vol. 43, Lille, France: PMLR, 2015, pp. 15–20. [Online]. Available: https://proceedings.mlr.press/v43/senft15.html (cit. on p. 77).

[84] K. Severinson-Eklundh, A. Green and H. Hüttenrauch, 'Social and collaborative aspects of interaction with a service robot,' in *Robotics and Autonomous Systems*, vol. 42, North-Holland, 2003, pp. 223–234. DOI: 10.1016/S0921-8890(02)00377-9 (cit. on pp. 8, 15).

[85] J. Shukla, J. Cristiano, J. Oliver and D. Puig, 'Robot Assisted Interventions for Individuals with Intellectual Disabilities: Impact on Users and Caregivers,' *International Journal of Social Robotics*, vol. 11, no. 4, pp. 631–649, 2019, ISSN: 18754805. DOI: 10.1007/s12369-019-00527-w. [Online]. Available:

https://link.springer.com/article/10.1007/s12369-019-00527-w (cit. on pp. 76, 104).

[86] C. L. Sidner, C. D. Kidd, C. Lee and N. Lesh, 'Where to look: A study of human-robot engagement,' in *Proceedings of the 9th international conference on Intelligent user interfaces*, ACM, 2004, pp. 78–84 (cit. on pp. 48–50).

[87] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh and C. Rich, 'Explorations in engagement for humans and robots,' *Artificial Intelligence*, vol. 166, no. 1-2, pp. 140–164, 2005 (cit. on pp. 49, 50, 76).

[88] C Spearman, 'The proof and measurement of association between two things,' *International Journal of Epidemiology*, vol. 39, no. 5, pp. 1137–1150, 2010, ISSN: 0300-5771. DOI: 10.1093/ije/dyq191. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21051364/ (cit. on p. 59).

[89] P. J. Standen, D. J. Brown, M. Taheri, M. J. Galvez Trigo, H. Boulton, A. Burton, M. J. Hallewell, J. G. Lathe, N. Shopland, M. A. Blanco Gonzalez, G. M. Kwiatkowska, E. Milli, S. Cobello, A. Mazzucato, M. Traversi and E. Hortal, 'An evaluation of an adaptive learning system based on multimodal affect recognition for learners with intellectual disabilities,' *British Journal of Educational Technology*, vol. 51, no. 5, pp. 1748–1765, 2020, ISSN: 14678535. DOI: 10.1111/bjet.13010. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/bjet.13010https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13010https://bera-journals.onlinelibrary.wiley.com/doi/10.1111/bjet.13010 (cit. on p. 76).

[90] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018 (cit. on p. 79).

[91] F. Tanaka, A. Cicourel and J. R. Movellan, 'Socialization between toddlers and robots at an early childhood education center,' *Proceedings of the National Academy of Sciences*, vol. 104, no. 46, pp. 17 954–17 958, 2007, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0707769104. (visited on 28th Sep. 2019) (cit. on pp. 51, 56, 57, 69, 101).

[92] A. Tanevska, F. Rea, G. Sandini, L. Cañamero and A. Sciutti, 'A Socially Adaptable Framework for Human-Robot Interaction,' *Frontiers in Robotics and AI*, vol. 7, p. 121, 2020, ISSN: 22969144. DOI: 10.3389/frobt.2020.00121. arXiv: 2003.11410. [Online]. Available: www.frontiersin.org (cit. on pp. 9, 77).

[93] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte *et al.*, 'Minerva: A second-generation museum tour-guide robot,' in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, IEEE, vol. 3, 1999 (cit. on pp. 15, 77).

[94] S. Tulli, D. A. Ambrossio, A. Najjar and F. J. Rodríguez Lera, 'Great expectations & Aborted business initiatives: The paradox of social robot between research and industry,' in *CEUR Workshop Proceedings*, vol. 2491, 2019 (cit. on p. 1).

[95] S. Wang and H. I. Christensen, '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*, 2018, pp. 158–165, ISBN: 9781538679807. DOI: 10.1109/ROMAN.2018.8525845. [Online]. Available: https://github.com/CogRob/ (cit. on p. 7).

[96] C. Webster and S. Ivanov, 'Public Perceptions of the Appropriateness of Robots in Museums and Galleries,' *Journal of Smart Tourism*, vol. 2, no. 1, pp. 33–39, 2022, ISSN: 2765-2157. DOI: 10.52255/SMARTTOURISM.2022.2.1.4. [Online]. Available: https://doi.org/10.52255/smarttourism.2022.2.1.4 (cit. on p. 16).

[97] J. Wiederer, A. Bouazizi, M. Troina, U. Kressel and V. Belagiannis, 'Anomaly detection in multi-agent trajectories for automated driving,' in *Proceedings of the 5th Conference on Robot Learning*, A. Faust, D. Hsu and G. Neumann, Eds., ser. Proceedings of Machine Learning Research, vol. 164, PMLR, 2022, pp. 1223–1233. [Online]. Available: https://proceedings.mlr.press/v164/wiederer22a.html (cit. on p. 103).

[98] H. Won Park, J. Busche, B. Schuller, C. Breazeal, R. W. Picard *et al.*, 'Personalized estimation of engagement from videos using active learning with deep reinforcement learning,' in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0 (cit. on pp. 51, 104).

[99] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, 'Aggregated residual transformations for deep neural networks,' in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500 (cit. on p. 62).

[100] V. A. Ziparo, L Iocchi, P. U. Lima, D Nardi and P. F. Palamara, 'Petri Net Plans: A framework for collaboration and coordination in multi-robot systems,' *Autonomous Agents and Multi-Agent Systems*, vol. 23, no. 3, pp. 344–383, 2011, ISSN: 13872532. DOI: 10.1007/s10458-010-9146-1. [Online]. Available: https://link.springer.com/article/10.1007/s10458-010-9146-1 (cit. on p. 19).