Maintaining, Monitoring, and Managing User Engagement in Social Robotics for Short-Term Interactions

Alan Lindsay[1], **Andrés A. Ramírez-Duque**[2], **Mary Ellen Foster**[2], **Ronald P. A. Petrick**[1]

[1]Heriot-Watt University, Edinburgh, UK
[2]University of Glasgow, Glasgow, UK
alan.lindsay@hw.ac.uk

Abstract

Social robotics has already demonstrated benefits across several areas, including robot-assisted education and healthcare. The use of plan-based approaches, which underpin the human-robot interaction with a planning model, are promising, especially in domains where collecting data in advance is challenging (e.g., medical domains). However, although the careful management of engagement during an interaction is critical for the success of social robots, the subject has not been fully explored in the context of plan-based solutions. In this paper we focus on a plan-based social robot system recently developed for use in a medical setting, and demonstrate how we have extended our robot system to maintain, monitor and manage real-time user engagement. We present an empirical evaluation where we sample the possible interactions that our system supports.

Introduction

Socially assistive robots (SARs) are embodied devices designed to interact with humans by communicating through mechanisms compatible with a human-centric approach (Feil-Seifer and Mataric 2005). The primary focus of SARs is to provide necessary aid to humans by engaging with them socially. SARs have proven helpful by assisting and supporting people experiencing stress or anxiety, such as children undergoing medical procedures (Trost et al. 2019; Moerman, van der Heide, and Heerink 2018; Dawe et al. 2019). Studies have compared short-term single-procedure exposure or long-term companionship (Ligthart, Neerincx, and Hindriks 2022), as well as the effectiveness of alternative robot-delivered interventions (Ali et al. 2019; Smakman et al. 2021; Rossi, Larafa, and Ruocco 2020).

In order that a social robot can have the intended positive impact on the user requires that the robot can initiate and maintain the user's engagement. In (de Haas, Vogt, and Krahmer 2021) they demonstrate that the selected strategy can impact on engagement and in (Szafir and Mutlu 2012; Del Duchetto and Hanheide 2022) they demonstrate the use of predicted engagement values on-line to alter the robot's behaviour. Within long-term interactions, reinforcement learning frameworks have been used to learn policies that maximise engagement (Del Duchetto and Hanheide 2022). However, existing work in social robotics for short-term interactions has typically used predicted engagement values in order to adapt scripted behaviour (e.g., increase volume of delivery), or interject re-engagement behaviours (Szafir and Mutlu 2012), limiting the scope of the strategies investigated.

In this work we consider short-term interactions between a robot and a child user. We introduce an existing system, which adopts a plan-based approach to generating interactions for supporting the child user within a medical setting. These interactions are specialised for the specific child, and their medical pathway. We then extend this planning model, and system in order to support the management of user engagement. The starting point for this is an existing model of engagement (O'Brien and Toms 2008), which identifies three stages: point of engagement, sustained engagement, and disengagement. We ground their model of engagement in our scenario and incorporate it into the existing interaction planning model. We then present our framework, which monitors user engagement, and updates the planning model to reflect its prediction of the user's real-time engagement, allowing the robot's strategy to be updated. In the evaluation we simulate patient pathways, and examine how the system updates the interaction in response to the simulated user's engagement. The generated interactions follow the model of engagement, and demonstrate changes in strategy, appropriate behaviour selection, and changes to the overall aims of the interaction.

Background

In this section we introduce the specific medical scenario that we consider in this work, and then define the planning formalism that use for specifying the robot's behaviour.

A Companion Robot for a Medical Procedure

In this work we focus on a particular medical setting, which involves supporting children during a painful and distressing medical procedure. In the specific clinical scenarios that we are targeting, the robot is placed in a small room together with the patient, along with one or more carers and a Health Care Provider (HCP) during the course of a single clinical procedure. The intravenous insertion (IVI) was identified as an appropriate procedure: This is one of the most commonly performed procedures in the context of children seeking medical care, and also one that can be painful and distressing for the child and for their parents or caregivers.

It has been shown that a companion robot can be effective at reducing the distress caused by IVI (Ali et al. 2019).

Planning Model

In our approach, the robot's behaviours are underpinned by a planning model, which uses a declarative representation to represent the domain knowledge and possible interactions concisely. We use a fully observable non-deterministic (FOND) planning model based on (Muise, McIlraith, and Beck 2012), which can be defined as a tuple $\langle \mathcal{F}, \mathcal{I}, \mathcal{G}, \mathcal{A} \rangle$, with fluents $\mathcal{F}$, initial state $\mathcal{I}$ (a full assignment to $\mathcal{F}$), a partial goal state $\mathcal{G}$, and a set of actions $\mathcal{A}$. Each action $a \in \mathcal{A}$ is a pair $\langle pre_a, \mathit{eff}_a \rangle$, with a precondition pre_a (a subset of $\mathcal{F}$ that must hold) and an effect eff_a (a set of possible outcomes—fluents that are made true or false). If an action defines one outcome, it is a deterministic action; otherwise, it is a non-deterministic action. Each action application results in an outcome, but the outcome cannot be chosen by the planner. A solution to the problem is a branched plan π, which describes the sequence of actions that will achieve the goal, given any outcome.

Related Work

Technological systems based on SARs (Feil-Seifer and Mataric 2005) provide unique opportunities to establish new mechanisms that use human-like social communication as a means to generate embodied interaction, with reported benefits in various domains, such as social, behavioural, physical, and cognitive well-being in different populations (Amirova et al. 2021; Henschel, Laban, and Cross 2021), in applications such as robot-assisted education (Johal 2020), autism diagnosis and therapy (Scassellati et al. 2018; Gomez Esteban et al. 2017),and Alzheimer therapy and elderly care (Tapus, Tapus, and Mataric 2009; Wada et al. 2004). In our case study a SAR is used in a paediatric healthcare settings to help alleviate children's distress and pain.

The use of planning to support interaction has a long history, and planning techniques have been applied previously in a range of social robots and interactive systems. Recent examples include (Waldhart, Gharbi, and Alami 2016; Sanelli et al. 2017; Kominis and Geffner 2017; Papaioannou, Dondrup, and Lemon 2018). The most similar system is the JAMES social robot bartender (Petrick and Foster 2013, 2020), which directly used an automated planner to choose the robot's physical, sensing, and interactive actions.

There is not a single shared interpretation of engagement in HRI. Attention is typically included (Sidner et al. 2004), and in some approaches, positive affective response is also included, e.g., (Poggi 2007). In this work, in order to separate the child's engagement with the interaction with potential affective response to the medical procedure, we relate engagement with attention (Sidner et al. 2004).

In (O'Brien and Toms 2008) they present a three stage model of engagement: point of engagement, sustaining engagement, and disengagement. This model is consistent with work in HRI with children (Brown and Howard 2013; Leite et al. 2016), and is the model that we adopt here. There have been different approaches to sustaining and regaining engagement (Szafir and Mutlu 2012; Cao et al. 2019). In (Szafir and Mutlu 2012), attention is predicted in real time, if the user's engagement reduces the robot modifies its communication style (e.g., increasing volume, and using more gestures). In (Cao et al. 2019) they identify three levels of disengagement and use different behaviours for each level. In the context of children, specific strategies have been tried (Brown and Howard 2013; Leite et al. 2016). In (Brown and Howard 2013) they compare verbal and non-verbal re-engagement strategies and show that verbal strategies are more effective. In (Leite et al. 2016) they use direct reference to the child's disengagement (e.g., 'Can you please pay attention?'), but do not observe much benefit. In long-term interactions it has been demonstrated that reinforcement learning can be effective at improving user engagement, e.g., (Del Duchetto and Hanheide 2022). In our work we consider interactions that are short-term, and where inappropriate choice of actions may cause or increase user discomfort and distress, limiting the applicability of approaches based on reinforcement learning.

System Overview

We have developed a fully functioning companion robot for operating in this scenario, which was designed using both a co-design (involving several stages and children, parents and HCPs) and targeted meetings between the technical team and the HCPs (Lindsay et al. 2022; Foster et al. 2023; Lindsay et al. 2024; Ramírez-Duque et al. 2024). We identified several main stages: introduction, preprocedure (optional site-check), procedure, debrief, and conclusion. The robot positions itself as a friendly and supportive companion, setting out positive expectations, and can present various supportive behaviours, including providing diversions and humour, practising coping strategies, role modelling, and providing positive reinforcements.

Our system architecture is composed of several components, including social signal processing, an interaction manager, a planning system, and a robot platform. The target robot platform is the SoftBank NAO, which is a humanoid robot with 25 degrees of freedom, which enables it to move and perform a large variety of actions. Additionally, NAO is equipped with a speaker, allowing the generation of different stimuli using multiple communication channels, for example, using verbal language such as speech and body language through gestures.

The low-level face analysis behaviour module is responsible for detecting the patient's face, identifying facial landmarks, head pose, gaze direction, and facial expression. Based on the above facial features, the social signal processing module estimates the current focus of attention and the head movement speed. This information is used to estimate the patient's emotional state, providing an indirect measure of affective states such as anxiety, valence, arousal, and engagement which are needed to control system behaviour.

At the centre of the architecture is the interaction manager, which ensures synchronised transitions between the internal states of the system/robot. The interaction manager integrates the information from the social signal components to estimate the user's affective state. It also makes requests

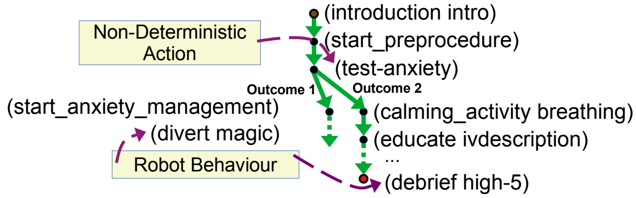

Figure 1: A partial plan showing actions (e.g., breathing exercises and high five), sensing actions (e.g., testing patient anxiety) and procedure actions (e.g., the preprocedure start).

```
(:action  pp_diverting_behaviour
  :parameters (?a - act ?c - category)
  :precondition (and (duringpreprocedure)
    (is_category ?a ?c) (= ?c diverting)
    (not (amperforminganxietymanagement))
  :effect (and (has_diverted)
    (done_activity ?a))
```

Figure 2: PDDL representation of the `pp_diverting_behaviour` action.

of the planning module, which is used during the interaction to determine the next action based on the current state and the goal. The planning system is built around the PRP planning system (Muise, McIlraith, and Beck 2012), which supports fully observable non-deterministic (FOND) planning models (Muise, McIlraith, and Beck 2012). We use a planning model to manage the interaction (described below). Finally, the social stimuli module interprets high-level actions and generates specific signals for each communication channel, whether through synthesised speech or non-verbal communication through gestures and body language.

A Companion Robot Planning Model

Our planning model is designed to capture child robot interactions for our medical setting (see [omitted] for more details). In our approach, propositional fluents model the situation in the room and state of the procedure, the patient's knowledge state (e.g., the robot has given certain information about the procedure), the patient's affective state (e.g., anxiety of the patient), and the progress of the interaction. The actions in the model can be separated into four groups: robot behaviour, procedure updates, implicit signals and explicit queries. For example, Figure 2 presents a PDDL action representation for a basic diverting robot behaviour for use during the preprocedure. The robot can perform a range of actions, including: distracting actions (e.g., dancing) and calming and instructive actions (e.g., stepping through breathing exercises); sensing actions for the medical scenario, e.g., to maintain the progress through the medical procedure; and patient focused sensing actions, e.g., to determine whether the patient is anxious in the interaction, each of which is represented in the planning model.

The main interaction captured in the planning model is structured along the possible patient pathways outlined with HCPs during the design process. A series of stages of the interaction were identified (e.g., introduction, pre-procedure, site-check, procedure, debrief, and conclusion), and the main variation within this sequence was determined (e.g., the length of stages like the procedure might vary considerably). These stages were used to organise the appropriate behaviours in each stage and in order to specify key objectives for the robot in each stage. For example, we can ensure that the robot delivers certain key information to the patient during the pre-procedure (e.g., regarding its role). The model represents the patient's current knowledge state, which can be used to select appropriate continuations of the interaction. Allowing the robot to progressively develop the child's current understanding of the procedure, coping strategies, and level of practice (based on cognitive-behavioural strategies used in (Jibb et al. 2018)). Plans generated for the model capture alternative interaction sequences based on input from the child's preferences and choices, variation in the medical pathway, and sensed valuation of the child's anxiety level.

Specialisation of the Interaction The robot's behaviour is specialised through the interaction with the patient, and the appropriate behaviour is selected using a variety of sensing actions. There is substantial variation in the medical procedure that can lead to variation in the length of the interaction (Lindsay et al. 2024). The robot also has a selection of questions that it can ask the user that will impact on the interaction, such as asking the user how it can best support them during the procedure – using calming, or diverting actions.

The robot's behaviour is also changed based on the patient's level of anxiety. An anxiety test sensing action (a sensing action that determines whether the patient's anxiety level is OK) determines whether the robot should use one of its interventions, designed to manage the user's affective sate. Figure 1 presents part of a branched plan, which includes the sensing action `sense_anxiety`. This allows the plan to capture strategies in alternative cases, e.g., either high (e.g., selecting an appropriate intervention) or normal anxiety (e.g., practising breathing exercises).

The aim in this paper is to use this model as a starting point, and to consider the issue of user engagement. Within the interaction the robot will provide support and companionship, and communicate important information regarding the procedure and possible strategies. However, to maximise the benefit, the interaction must be engaging for the user. The remainder of this work considers a framework that can use this planning model, extend it with a model of engagement, and use it to select behaviours for a robot that will attempt to maintain engagement with a child, while managing disengagement and re-engagement when necessary.

Extending an Interaction with an Engagement Model

In this work we focus on social engagement between the child patient and the robot (see (Oertel et al. 2020) for a review). We consider the model of engagement as a process proposed in (O'Brien and Toms 2008) and adopted for child-robot interactions (Brown and Howard 2013; Leite et al. 2016), which identifies three stages of engagement: point of engagement, period of engagement, and disengagement. For

```
(:action  point_of_engagement
  :precondition (not (engaged))
  :effect (engaged))

(:action  sustain_engagement
  :precondition (engaged))

(:action  disengagement
  :precondition (engaged)
  :effect (not (engaged)))
```

Figure 3: PDDL representations of the three actions in the engagement model presented in (O'Brien and Toms 2008).

```
(:action  pp_diverting_behaviour
  :parameters (?a - act ?c0 ?c1 - category)
  :precondition (and (engaged)
    (previous ?c0) (not (= ?c0 ?c1))
    (not (done_activity ?a)) (= ?c1 diverting)
    (is_category ?a ?c1) (duringpreprocedure)
    (not (amperforminganxietymanagement))
  :effect (and (previous ?c1)
    (not (previous ?c0)) (done_activity ?a)
    (has_diverted)))
```

Figure 4: PDDL representation of the pp_diverting_behaviour action, which extends the original model presented in Figure 2, with a simple model for sustaining engagement (in red).

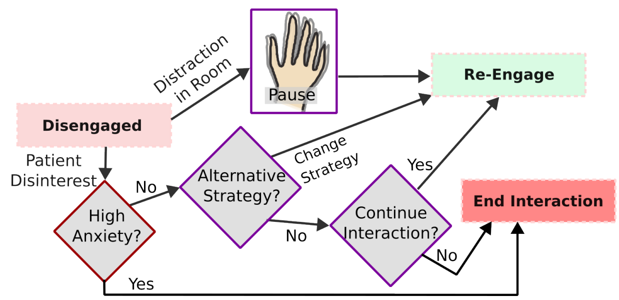

Figure 5: After disengagement the planning model captures the strategy for continuing the interaction. If the disengagement was as a result of a distraction, the system uses a pause action, before attempting to re-engage the user. In the case of disinterest, the system first attempts to change strategy, and if that fails then gives the option to end the interaction.

each of these stages they have identified attributes that can be associated with the stage. They also observe that these stages can repeat, and the user can become re-engaged. The attributes include both aspects of the robot's behaviour, and also aspects including its novelty, and its aesthetics. During a related co-design (Foster et al. 2023), several aspects of the robot's aesthetic were identified as very important, including its appearance (e.g., welcoming colour and no red eyes), delivery (e.g., nice sounding voice, with friendly gesturing), and interaction (e.g., age-appropriate; the robot should have a backstory). In this section we focus on the behaviour aspect of the engagement model (please refer to (Lindsay et al. 2024) for more detail about the overall system design). We use the model of engagement in (O'Brien and Toms 2008) with the information from the co-design (Foster et al. 2023), and the specification of the scenario, in order to characterise the model's three main stages in the context of our scenario. We first present a basic PDDL representation, which captures the three stages of the model. We then identify aspects of these stages that are important in our work, and extend the representation using specific situations in our scenario.

Model of Engagement

We can represent the three stage model of engagement in a planning model by using a representation of the user's engagement: a proposition that is True when the user is engaged. In this case we use the `engaged` proposition. In Figure 3 we present PDDL representations of the three stages. The `point_of_engagement` action has the precondition of not engaged and records the transition to engaged. The `sustain_engagement` action simply insists that the engaged proposition holds. And finally, the `disengagement` action transitions the engaged proposition back to False.

Sustaining Engagement

It is anticipated that the robot will lead to positive emotional response, with the child enjoying the interaction. In terms of maintaining engagement, this will require robot behaviour that keeps the child's attention and interest. In order that the children are interested requires that its behaviours are both age appropriate, varied, and provide ways for the child to influence its behaviours. Our baseline model already uses age appropriate behaviours, and provides opportunities to interact with the system. In this part we extend the baseline model to incorporate a simple model for sustaining engagement, to support the planner in making engaging interactions.

Incorporating a Basic Model of Sustaining Engagement

For the purpose of our short-term interaction we adopt a simple model for sustaining engagement. In particular, we promote novelty by i. preventing repeated robot behaviours, and ii. ensuring that adjacent behaviours are of different categories. We have extended the representations of the robot behaviours with the sustaining engagement model. For example, Figure 4 presents the extended PDDL for the `pp_diverting_behaviour` action (the original was presented in Figure 2). It incorporates the structure from the `sustain_engagement` action (Figure 3), and introduces a parameter, which represents a robot behaviour (activity), and the category of the activity. The precondition insists that the category is different from the previous category, and that the activity has not already been done. The effect makes the relevant updates to record that the activity is now done, and the current action is recorded as the previous action.

Disengagement

Our scenario is set in a typically busy emergency department, with continual distraction. Within the room the health care provider will be moving between stations in order to

```
(:action  end_interaction
  :parameters (?a - act)
  :precondition (and (engaged) (= ?a bowout)
    (high_anxiety) (during_preprocedure)
    (performing_anxiety_management))
  :effect (and (not (engaged))
    (done ?a) (interaction_ended)
    (not (performing_anxiety_management))))

(:action  pause_interaction
  :parameters (?a - act ?c - category)
  :precondition (and (engaged)
    (during_preprocedure) (= ?c info))
    (= ?a pause) (ok_anxiety) (previous ?c))
  :effect (and (not (engaged)) (doing_pause)
    (not (U_requires_proc_info))))
```

Figure 6: PDDL representation of the `end_interaction` and `pause_interaction` actions, which extends the `disengagement` action (Figure 3) for the cases of high and normal anxiety.

```
(:action  introduction
  :parameters (?a - activity)
  :precondition (and (not (engaged))
    (= ?a intro) (pre_interaction))
  :effect (and (engaged) (during_interaction)
    (not (pre_interaction))))

(:action query_activity_preference
  :parameters (?a1 ?a2 - act ?c1 ?c2 - cat)
  :precondition (and (not (engaged))
    (is_category ?a1 ?c1) (= ?c2 diverting)
    (is_category ?a2 ?c2) (not (= ?a1 ?a2)))
  :effect (and (engaged)
      (U_preference_set)
      (oneof (U_selected ?a1)
             (U_selected ?a2))))
```

Figure 7: PDDL representation of the `introduction` and `query_activity_preference` actions, which extend the `point_of_engagement` action (Figure 3) for the cases of the robot's introduction to initiate the interaction and using a user query to re-engage the child.

prepare for the procedure, and there might be traffic coming in and out of the room. We would therefore expect that at times the patient will become disengaged from the interaction. Disengagement can also be a result of negative affect, which in our scenario might be caused by boredom, and frustration with the technology. This might be because the robot is not behaving in a way that the child likes (e.g., they find the information boring), or they are not able to make the robot do something they want (e.g., lack of interaction).

In Figure 5 we present the categorisation of disengagement and subsequent strategy in our model. A typical case is where the user might just have been attracted to something else (distraction in room), the robot pauses its interaction for a brief interval. The `pause_interaction` action (bottom of Figure 6) demonstrates the extension of the `disengagement` action to implement a pause in the interaction. This allows a gap while the patient's attention is taken somewhere else. The robot will then continue with a re-engagement action (see below).

An alternative case is that patient becomes disinterested in the robot (patient disinterest). The first approach is to consider the context that led to the child becoming disengaged and to alter the strategy in order to respond appropriately (change strategy). In the co-design it became apparent that while some children are interested in receiving some information about the medical procedure, there are others that do not. We can therefore interpret the patient's disengagement during an information task as an opportunity in order to update the robot's strategy away from providing information, and towards other types of behaviour. As a consequence we reduce the number of the robot's information providing goals (see `pause_interaction` action in Figure 6).

In the current model there are certain special cases, where the interaction cannot proceed normally. These points have been established with HCPs as constraints on the space of patient pathways that will be considered during the future clinical trials. These include situations where there have been IVI procedure complications, potentially leading to an emergency situation and making further participation in the trial inappropriate. Alternatively, there are situations where robot intervention might fail (e.g., the patient may be disengaged, or their anxiety may continue to be high after intervention). A concern raised during co-design was that the robot does not become an additional noise in the room. In particular, in some case the patient will not engage with the robot. This can occur even if the patient is distressed. In these cases the robot should back off, otherwise it may cause further distress. The `end_interaction` action (Figure 6) presents one of these situations for situations with high patient anxiety and no engagement.

Point of Engagement

The first stage in the model of engagement is the point of engagement, which includes both the initial engagement, as well as possible re-engagement.

Initial Engagement The start of the interaction is very important for setting the expectations of the child. In the initial point of engagement the robot attempts to position itself as a companion for the procedure. It attempts to set up positive expectations for the child: "I am so excited to play with you today", and sets its role as friendly and supportive: "let's do this together". This provides an outline of how the robot can help the child during the interaction. The `introduction` action (see Figure 7) starts the interaction, and demonstrates the extension of the `point_of_engagement` action. The action introduces the activity parameter, which is associated with the robot's introduction behaviour.

Re-engaging User After Disengagement Disengagement is a common part of any interaction, however, previous research has shown that overall engagement can be improved through the appropriate use of re-engagement strategies (Cao et al. 2019). Our primary re-engagement strategy

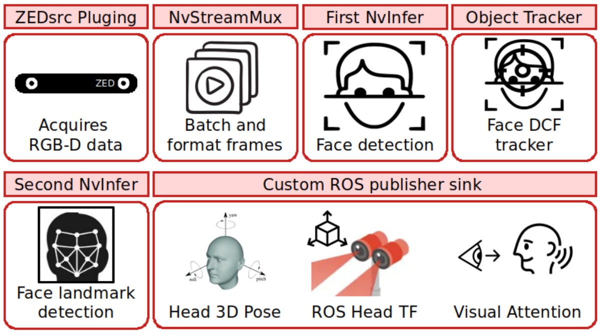

Figure 8: Low-level face analysis pipeline using Nvidia Deepstream framework.

involves asking the child their preference on the next activity. This is because interaction can lead to better engagement with children (Ligthart, Neerincx, and Hindriks 2020). In this case the robot will offer a choice of alternative next behaviours ('would you like me to dance or sing next?'). This is captured by the `query_activity_preference` action (see Figure 7 bottom). The action defines two outcomes: a preference for activity one (e.g., dancing), or a preference for activity two (e.g., singing). In this case we force that at least one activity is a diverting activity. The next steps are to implement the suggested behaviour.

The selection of the appropriate re-engagement actions is made by the planner, which allows it to make the choice in a context sensitive manner. For example, during the preprocess, the user can select between different behaviours, including diverting and calming. However, during the sitecheck the patient must be still, and the robot offers only calming actions. Another example is while the HCP is carrying out the actual procedure, where the robot cannot rely on direct responses from the child. In this case the robot instead uses an alternative strategy, where a highly diverting actions is selected. In these cases, the selection of user options, and appropriate behaviours is backed by the planning model, which ensures that the choices are appropriate for the context, and that there are appropriate behaviours for the remainder of the interaction.

Monitoring User Engagement

One of our system's key features is its ability to select the appropriate behaviour based on the user's level of engagement with the robot. This is achieved within a pipeline that aims to predict the patient's visual focus of attention. Due to the physical constraints of the robot deployment, this was limited to the patient's head analysis – It has been demonstrated that head orientation and eye-gaze direction can be used to estimate a child's engagement with a robot (de Haas, Vogt, and Krahmer 2021). In fact, predicting the patient's visual focus of attention automatically in this scenario proves to be challenging due to various limiting factors: 1.) There is limited space near the patient, and constant staff movement causes occlusion. 2.) The patient is likely to be wearing a surgical mask. 3.) The system must be portable and mobile; it must be able to move between the different emergency rooms with agility, which reduces the possibility of using fixed cameras and Internet connections via LAN and WLAN due to interference.

The automatic facial analysis pipeline is based on Nvidia DeepStream SDK and was deployed using a Jetson Nano board. During a practical application, the head position and 3D orientation, visual attention and the speed of movement of the patient's head are estimated. We used the FaceX-Zoo framework in the face and landmarks detection stage (Wang et al. 2021). We selected two models, a PyTorch implementation of the RetinaFace model (Deng et al. 2019) and the Practical Facial Landmark Detector (PFLD) (Guo et al. 2019). These models were retrained using the MegaFace-Mask database, improving detection in images of subjects while wearing a mask.

Taking advantage of ready-to-use hardware-accelerated plugins, we used TensorRT, NVIDIA's inference accelerator runtime, for model inference. In addition, we used built-in plugins, including the Nvidia-adapted Discriminative Correlation Filter (DCF) tracker and the GazeNet inference. GazeNet detects the patient's gaze vector and point of regard, and it was trained on an Nvidia proprietary dataset.

As a final element of the pipeline, our system uses the Point Distribution Model (PDM) from OpenFace toolkit (Baltrusaitis et al. 2018) to calculate the head 3D pose and a ROS-based plugin to represent the position of the patient's head as a coordinate system defined with homogeneous transformation respect to the camera's optical coordinate system. To calculate when the patient focuses on the robot in each frame, we simulate the patient's field of vision as a cone with a 30-degree opening and two meters depth. We associate a coordinate system to the robot and place it approximately in the middle of it (NAO chest button). Using the functionality of the tf library in ROS, we calculated the relative transformation between the robot's coordinate system and the patient's head. If it was within the field of vision, we classified the frame with a Boolean valuation *1: attended the robot*; otherwise, *0: did not attend the robot*. Finally, we publish the features estimated along the pipeline. A diagram that summarises all the gst-plugins implemented in the pipeline is shown in Figure 8. Each block represents a specific plugin, and together, they are optimised through memory management with zero-memory copying between plugins, ensuring its performance.

Analysing Sensed Data

We adopt a simple thresholding method to interpret the child's predicted visual focus of attention data for use in our planning model. At each time point, we examine the previous ten data frames (constituting a 1-second period) and tagged this as attended the robot or did not attend the robot as explained above. These values are maintained in a sliding window, which retains a 60-second view of the user's engagement. Every time that the system requests an assessment of the user's engagement, these values are averaged and compared to a thresholding value, *threshold* (we have used *threshold*$=$ 0.5 in an associated study (Lindsay et al. 2024)).

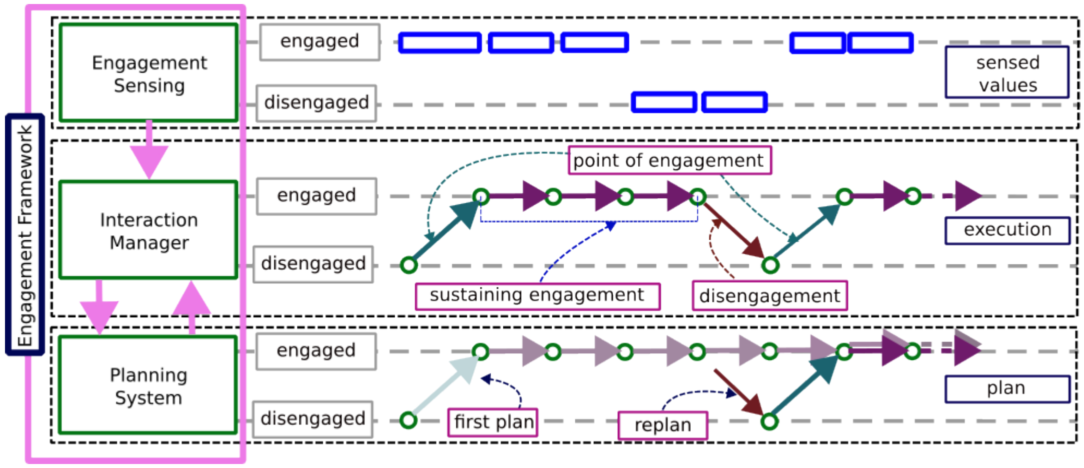

Figure 9: The left shows the main components of the engagement framework. In our framework the planner generates sequences intended to sustain engagement (bottom). The framework uses sensed data from the user (top) to force appropriate disengagement actions to appropriately update the planning model state, driving new sequence generation (middle).

A Framework for Managing Engagement

In this section we present a framework for using the engagement model and sensing presented in the previous sections. The framework (see the left side of Figure 9) uses a planner (planning system) to generate sequences of actions that are intended to both sustain engagement (based on our model) and generate interactions that include user customisation, alternative medical pathways, and monitoring and management of the user's affective state (see the 'Plan-Based Interaction' Section). Around the planner the framework observes the user's engagement level (engagement sensing), and if the user becomes disengaged it uses the planner to generate a new plan, which must select an appropriate `disengagement` action.

As part of the framework we have the planning model of the scenario, which includes point of engagement, sustaining, and disengagement actions (see the 'Model of Engagement' Section), as well as additional actions overviewed in the 'Plan-Based Interaction' Section (e.g., anxiety tests, medical pathway updates). Given an initial situation, where the child has not yet engaged with the robot, the planning model will be used to generate a plan. This plan will initiate the interaction (see Figure 7) and generate a branched plan for the interaction based on an assumption of sustained engagement (bottom of Figure 9).

The framework will step through this plan incrementally, using appropriate sensing capabilities to elicit user preferences, sense the user's affective state, and progress its model of the medical procedure. After each action it will also monitor the user's engagement level (top of Figure 9). If the user is continuing to be engaged by the robot then the system will continue executing the plan (selecting appropriate branches as appropriate) until the goal (end of the interaction). However, due to the medical scenario involving likely distractions, and variation in children's interests, it is possible that during the interaction, the child will become disengaged.

If the child becomes disengaged the system will update the planning model by using a special proposition, which cannot be added by the planner. The special proposition prevents robot behaviour actions, and is removed by the application of an appropriate disengagement action. This forces the planner to appropriately register the disengagement, applying the appropriate disengage action for the context. For example, we presented in Figure 6 two alternative extensions of the `disengagement` action (e.g., pausing, changing strategy, or ending the interaction). After executing the appropriate action, the system then requests the next action from the planner, which will typically aim to re-engage the user (using a point of engagement action, e.g., Figure 7), and continue with the interaction (potentially with a new strategy). The middle of Figure 9 shows how the framework uses replanning after disengagement in order to allow the system to adapt the interaction to the user's level of engagement.

Evaluation

In this section, we use simulated responses to examine the interactions generated by the engagement framework, presented in Figure 9, and implemented within the interaction manager of a functioning robot system (Ramírez-Duque et al. 2024). The system is modular, and we have created dummy components for the robot behaviour implementer, the sensing, and the web-browser components. In response to choice queries (e.g., preference between activities or the current anxiety value), the dummy web-server and sensing components were specified to return a random choice between the alternatives (e.g., Bruno-dance or OK-anxiety), allowing us to test the integration of the presented framework with the robot's planning system.

We sampled 100 simulated interactions. The dummy components selected a random value for each possible decision, allowing us to represent interactions for alternative patient personalisations, choices and anxiety levels. We added a random chance that the system should record disengagement (simulating user disengagement): After each action,

# Actions			Planning time			Points of Eng'			Distance		
Avg.	Max.	Min.	Avg.	Max.	Min.	Avg.	Max.	Min.	Avg.	Max.	Min.
24.56	37	11	7.65	13.71	3.77	1.87	5	1	15.99	31	0

Table 1: Summary of the 100 simulated interactions, reporting average, min and max, for the number of actions (# actions), planning time, points of engagement (this includes first engagement), and a measure of sequence distance (see text).

there is a 1 in 15 chance that disengagement is registered. This allowed us to explore how the system responded to disengagement in different contexts and monitor its decision to re-engage or end the interaction.

Table 1 presents a summary of the simulated interactions. The system continued each interaction and brought the intervention to an appropriate conclusion. In 70% of the cases, the generated sequence represented a successful interaction that proceeded through the entire procedure. In each other case, the interaction was brought to an early conclusion because the sequence had gone out of scope, e.g., complications in the procedure.

The results show that there is a wide range in lengths of execution sequences (11-37 steps). This is largely due to the variation in the underlying medical procedure. For example, in some cases, the HCP can locate a promising site immediately, and the procedure is straightforward. In other situations, both the site-check and the procedure can be prolonged. Of course, there are also special cases where the robot drops out. For example, the 11-step plan involves the interaction being brought to an early end.

The planning time is, on average, 0.31 seconds per step. Our architecture asks the planner for an action at each step. In the first instance, the planner generates a state action policy. In subsequent states it first attempts to lookup the state in the current policy. If it fails, it will generate a new policy. In most cases, the branched plan will capture the plan in each case. This ensures that the plan is balanced enough to satisfy each predictable choice point. In the case of disengagement, the planner will typically be forced to replan. There is an increase in average planning time per action as the number of points of engagement increases (e.g., from 0.29 for 1 point of engagement to 0.34 for 5 points of engagement). However, even when forced to replan, the planning time is reasonable for the application.

It is expected that due to the busy nature of the environment, the robot may have to re-engage with the patient a number of times during the interaction. However, we anticipate this will not typically be more than one or two times (i.e., 1 to 3 points of engagement). The results in Table 1 cover interactions with 1 to 5 points of engagement. In each case, the system responded appropriately, registering the signal, updating the state, and, consequently, forcing the planner to select an appropriate disengagement action. In nearly all of the interactions, a re-engagement action was selected, and the interaction continued (in some cases with an updated strategy). In the other case, the patient exhibited high anxiety and was disengaged.

Finally, we compared the plans to examine their similarity. To do this we used the Levenshtein distance (Levenshtein 1966): the distance between two-word sequences which provides a measure of the edit difference between the sequences while also respecting order. In our case, we use unique words for each ground action. This measure was used as it has been demonstrated that it provides a measure for comparing plan similarity (Coman and Munoz-Avila 2011). The results demonstrate that the sequences of actions generated by the system are relatively varied (i.e., the average number of differences is more than half the average sequence length), demonstrating the impact of personalisation on the generated interactions. Two plans were repeated in the results, and they were both very short (12 and 13 steps).

Conclusion and Future Work

In this work, we have considered extending a system capable of supporting a plan-based interaction to manage user engagement. We have used a model of engagement that includes point of engagement, sustaining engagement, disengagement, and re-engagement to extend a planning model for a companion robot interaction. Through using the planning model, we can rely on the planning model's state representation, allowing the selection of context-appropriate disengagement and re-engagement actions. We have demonstrated how the context can be exploited in order to update the system's strategy in response to the observed behaviour. We have presented the sensing system that we will use in order to monitor the user and trigger disengagement events in the system. We have presented the results of 100 simulated interactions, where our framework was used to create the appropriate responses based on the simulated user responses. This work forms part of a robot system that is currently undergoing a usability study prior to a clinical trial.

Acknowledgements

The authors acknowledge the SSHRC-UKRI Canada-UK Artificial Intelligence Initiative (UKRI grant ES/T01296/1) for financial support of this project. We would also like to acknowledge all patient partners, research and clinical staff, and youth and families who have made this project possible.

References

Ali, S.; Manaloor, R.; Ma, K.; Sivakumar, M.; Vandermeer, B.; Beran, T.; Scott, S.; Graham, T.; Curtis, S.; Jou, H.; et al. 2019. LO63: humanoid robot-based distraction to reduce pain and distress during venipuncture in the pediatric emergency department: a randomized controlled trial. *Canadian Journal of Emergency Medicine*.

Amirova, A.; Rakhymbayeva, N.; Yadollahi, E.; Sandygulova, A.; and Johal, W. 2021. 10 Years of Human-NAO Interaction Research: A Scoping Review. *Frontiers in Robotics and AI*, 8.

Baltrusaitis, T.; Zadeh, A.; Lim, Y. C.; and Morency, L. P. 2018. OpenFace 2.0: Facial behavior analysis toolkit. *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, 59–66.

Brown, L.; and Howard, A. M. 2013. Engaging children in math education using a socially interactive humanoid robot. In *IEEE-RAS International Conference on Humanoid Robots*.

Cao, H.-L.; Torrico Moron, P. C.; Esteban, P. G.; De Beir, A.; Bagheri, E.; Lefeber, D.; and Vanderborght, B. 2019. "Hmm, Did You Hear What I Just Said?": Development of a Re-Engagement System for Socially Interactive Robots. *Robotics*, 8(4).

Coman, A.; and Munoz-Avila, H. 2011. Generating Diverse Plans Using Quantitative and Qualitative Plan Distance Metrics. In *AAAI*, 946–951. Citeseer.

Dawe, J.; Sutherland, C.; Barco, A.; and Broadbent, E. 2019. Can social robots help children in healthcare contexts? A scoping review. *BMJ Paediatrics Open*, 3(1).

de Haas, M.; Vogt, P.; and Krahmer, E. 2021. When Preschoolers Interact with an Educational Robot, Does Robot Feedback Influence Engagement? *Multimodal Technologies and Interaction*, 5(12).

Del Duchetto, F.; and Hanheide, M. 2022. Learning on the Job: Long-Term Behavioural Adaptation in Human-Robot Interactions. *IEEE Robotics and Automation Letters*, 7(3): 6934–6941.

Deng, J.; Guo, J.; Zhou, Y.; Yu, J.; Kotsia, I.; and Zafeiriou, S. 2019. RetinaFace: Single-stage Dense Face Localisation in the Wild.

Feil-Seifer, D.; and Mataric, M. 2005. Socially Assistive Robotics. In *Proceedings of the International Conference on Rehabilitation Robotics*.

Foster, M. E.; Candelaria, P.; Dwyer, L. J.; Hudson, S.; Lindsay, A.; Nishat, F.; Pacquing, M.; Petrick, R. P. A.; Ramírez-Duque, A. A.; Stinson, J.; Zeller, F.; and Ali, S. 2023. Co-Design of a Social Robot for Distraction in the Paediatric Emergency Department. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery.

Gomez Esteban, P.; Baxter, P. E.; Belpaeme, T.; Billing, E.; Cai, H.; Cao, H.-L.; Coeckelbergh, M.; Costescu, C.; David, D.; De Beir, A.; Fang, Y.; Ju, Z.; Kennedy, J.; et al. 2017. How to Build a Supervised Autonomous System for Robot-Enhanced Therapy for Children with Autism Spectrum Disorder. *Paladyn Journal of Behavioral Robotics*, 8(1): 18–38.

Guo, X.; Li, S.; Yu, J.; Zhang, J.; Ma, J.; Ma, L.; Liu, W.; and Ling, H. 2019. PFLD: A Practical Facial Landmark Detector.

Henschel, A.; Laban, G.; and Cross, E. S. 2021. What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You. *Current Robotics Reports*, 2: 9–19.

Jibb, L. A.; Birnie, K. A.; Nathan, P. C.; Beran, T. N.; Hum, V.; Victor, J. C.; and Stinson, J. N. 2018. Using the MEDiPORT humanoid robot to reduce procedural pain and distress in children with cancer: A pilot randomized controlled trial. *Pediatric Blood & Cancer*, 65(9): e27242.

Johal, W. 2020. Research Trends in Social Robots for Learning. *Current Robotics Reports*, 1: 75–83.

Kominis, F.; and Geffner, H. 2017. Multiagent online planning with nested beliefs and dialogue. In *Proceedings of the International Conference on Automated Planning and Scheduling*.

Leite, I.; McCoy, M.; Lohani, M.; Salomons, N.; McElvaine, K.; Stokes, C.; Rivers, S.; and Scassellati, B. 2016. Autonomous disengagement classification and repair in multiparty child-robot interaction. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 525–532. IEEE.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions and reversals. *Cybernetics and Control Theory*, 10: 707–710.

Ligthart, M. E.; Neerincx, M. A.; and Hindriks, K. V. 2020. Design Patterns for an Interactive Storytelling Robot to Support Children's Engagement and Agency. Association for Computing Machinery.

Ligthart, M. E.; Neerincx, M. A.; and Hindriks, K. V. 2022. Memory-Based Personalization for Fostering a Long-Term Child-Robot Relationship. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*.

Lindsay, A.; Ramírez-Duque, A. A.; Petrick, R. P. A.; and Foster, M. E. 2022. A Socially Assistive Robot using Automated Planning in a Paediatric Clinical Setting. In *Proceedings of the AAAI Fall Symposium on AI for Human-Robot Interaction*.

Lindsay, A.; Ramírez-Duque, A. A.; Petrick, R. P. A.; and Foster, M. E. 2024. A Socially Assistive Robot using Automated Planning in a Paediatric Clinical Setting. In *Proceedings of the International Symposium on Technological Advances in HRI*.

Moerman, C. J.; van der Heide, L.; and Heerink, M. 2018. Social robots to support children's well-being under medical treatment: A systematic state-of-the-art review. *Journal of Child Health Care*, 23: 596–612.

Muise, C.; McIlraith, S.; and Beck, C. 2012. Improved non-deterministic planning by exploiting state relevance. In *Proceedings of the International Conference on Automated Planning and Scheduling*, 172–180.

O'Brien, H. L.; and Toms, E. G. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6): 938–955.

Oertel, C.; Castellano, G.; Chetouani, M.; Nasir, J.; Obaid, M.; Pelachaud, C.; and Peters, C. 2020. Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 7: 92.

Papaioannou, I.; Dondrup, C.; and Lemon, O. 2018. Human-robot interaction requires more than slot filling-multi-threaded dialogue for collaborative tasks and social conversation. In *FAIM/ISCA Workshop on Artificial Intelligence for Multimodal HRI*.

Petrick, R.; and Foster, M. E. 2020. Knowledge engineering and planning for social human–robot interaction: A case study. In *Knowledge Engineering Tools and Techniques for AI Planning*.

Petrick, R. P.; and Foster, M. E. 2013. Planning for Social Interaction in a Robot Bartender Domain. In *Proceedings of the International Conference on Automated Planning and Scheduling*.

Poggi, I. 2007. *Mind, hands, face and body: a goal and belief view of multimodal communication*. Weidler.

Ramírez-Duque, A. A.; Lindsay, A.; Foster, M. E.; and Petrick, R. P. 2024. A Lightweight Artificial Cognition Model for Socio-Affective Human-Robot Interaction. In *Proceedings of the ACM/IEEE International Conference on HRI*.

Rossi, S.; Larafa, M.; and Ruocco, M. 2020. Emotional and Behavioural Distraction by a Social Robot for Children Anxiety Reduction During Vaccination. *International Journal of Social Robotics*, 12(3): 765–777.

Sanelli, V.; Cashmore, M.; Magazzeni, D.; and Iocchi, L. 2017. Short-Term Human-Robot Interaction through Conditional Planning and Execution. In *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS)*, 540–548.

Scassellati, B.; Boccanfuso, L.; Huang, C.-M.; Mademtzi, M.; Qin, M.; Salomons, N.; Ventola, P.; and Shic, F. 2018. Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*, 3(21): eaat7544.

Sidner, C. L.; Kidd, C. D.; Lee, C.; and Lesh, N. 2004. Where to look: a study of human-robot engagement. In *Proceedings of the 9th international conference on Intelligent user interfaces*, 78–84.

Smakman, M. H. J.; Smit, K.; Buser, L.; Monshouwer, T.; van Putten, N.; Trip, T.; Schoof, C.; Preciado, D. F.; Konijn, E. A.; van der Roest, E. M.; and Tiel Groenestege, W. M. 2021. Mitigating Children's Pain and Anxiety during Blood Draw Using Social Robots. *Electronics*, 10(10): 1211.

Szafir, D.; and Mutlu, B. 2012. Pay attention! designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Tapus, A.; Tapus, C.; and Mataric, M. J. 2009. The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In *Proceedings of the IEEE International Conference on Rehabilitation Robotics*.

Trost, M. J.; Ford, A. R.; Kysh, L.; Gold, J. I.; and Matarić, M. 2019. Socially assistive robots for helping pediatric distress and pain: a review of current evidence and recommendations for future research and practice. *The Clinical journal of pain*.

Wada, K.; Shibata, T.; Saito, T.; and Tanie, K. 2004. Effects of robot-assisted activity for elderly people and nurses at a day service center. *Proceedings of the IEEE*.

Waldhart, J.; Gharbi, M.; and Alami, R. 2016. A novel software combining task and motion planning for human-robot interaction. In *2016 AAAI Fall Symposium Series*.

Wang, J.; Liu, Y.; Hu, Y.; Shi, H.; and Mei, T. 2021. FaceX-Zoo: A PyTorch Toolbox for Face Recognition.