Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Master Thesis Proposal

Deep Learning-Based Personalisation of Robot Behaviour for Assistive Robotics

Michał Stolarz

Supervised by

Prof. Dr. Paul G. Plöger

MSc. Alex Mitrevski

January 2023

1 Introduction[1]

1.1 Motivation

One of the objectives of robot-assisted therapy (RAT) [27] is increasing the autonomy of the robot that is used during therapy sessions; this has the purpose of reducing the necessary therapist interactions with the robot, while still keeping the therapist in control of the sessions at all times. In the context of RAT, robot programs are usually developed in such a way that they can be used generically for different individuals; however, individuals may have different reactions to specific stimuli and, depending on their concrete needs, may also benefit from therapy sessions focusing on specific aspects. This means that a generic RAT approach may not be optimal for effective treatment of individuals; instead, the robot should be able to adapt its behaviour to the needs of each individual and therapy session [27, 73, 78].

The motivation behind this work is a therapy for children with Autism Spectrum Disorder (ASD). This problem is of a big relevance as in the European Union, there are over 5 million people affected by autism [1] and it is estimated that 1 in 160 children all over the world is diagnosed with ASD [36]. People with ASD often have difficulties in social interaction and communication. To alleviate the effects of ASD, individualized therapies are provided. However, autistic children find robots easier to communicate with than humans [66], thus Robot-Assisted Therapies (RATs) have been being investigated. During RAT, most of the time therapists have to control the robot remotely (Wizard of Oz approach) [1] [22, 50, 67, 72]. Because of it, the therapist might not be able to fully focus on the therapy and react appropriately to the child's behaviour [9]. To reduce their workload, the autonomy of the robot has to be increased, namely it should be able to interpret a child's behaviour and adapt its actions to the individual needs of the child [27].

1.2 Adaptation Techniques

Adaptation is possible if the robot actively learns a user model that encodes certain attributes of the user. The user model can be integrated into a robot

[1] Parts of this chapter have been published in [89, 91]

decision-making algorithm [71] called a behaviour model, which allows the system to choose appropriate robot reactions in response to the actions of each individual user. Personalisation refers to the adaptation of the system to the individual user over time [71] and can be solved by using Interactive Machine Learning (IML), which involves the user in the learning loop [83]. IML usually makes use of *learning from guidance* or *learning from feedback*. Learning from guidance relies on an external supervisor (e.g. therapist), who provides expert knowledge to the system. The supervisor is able to assess the decisions of the robot before being executed, namely they are able to accept, or alternatively reject and override the suggested reaction of the robot. This solution guarantees that the system will not execute any undesirable actions during learning, but is sensitive to the mistakes of the supervising person. On the other hand, learning from feedback uses direct feedback from the user (e.g. engagement level of the user). As there is no supervising person, the robot has to explore by itself what effects its actions have.

1.3 Project Goal

The main problem during RAT for children with autism is that the therapists have to control the robot manually, which might meaningfully increase their workload. This means that there is a need for a personalised behaviour model which will increase the autonomy of the robot. The model should interpret and continuously adapt to the behaviour of the individual child under therapy, as each child might have different ASD symptoms. The therapy for children with ASD usually consists of games designed by the therapists. This means that the developed learning algorithm should be able to enable the robot to personalise the difficulty of the game activities to the individual child's skill level. Additionally, the robot should also react appropriately when interacting with the child does not go as planned. That means that the robot should prevent them from getting bored, disengaged or demotivated, by executing actions such as giving verbal motivating feedback or simple motions (e.g. waving gesture) that would draw the child's attention back to the game.

Currently, in order to enable the robot to react appropriately in various social situation during an interaction with the user, many works make use of an engage-

ment estimator. Engagement is the feature that is often used for the development of behaviour models [23, 80, 81, 95]. It can be measured with the use of EEG headset [95], but an external engagement observer might be more convenient for children with ASD, as they may be overwhelmed by the sensory stimuli if they need to wear an additional device during therapy [37]. In the literature, several types of algorithms that estimate engagement from features obtained using the OpenFace library [4, 36, 39, 40], eye gaze [41], body posture [65] or visual data [24, 49] can be found. Some of these are also able to capture and classify temporal data [24, 39].

However, the model estimating an engagement used further for decision-making process is not perfect and might introduce an additional error to the behaviour model. This is a problem of a significant impact on the robot behaviour and was mentioned in our recent work [89]. To alleviate the impact of false predictions on the feature level we suggest to turn towards data-driven methods that will be able to use a raw sensor data instead of high-level features (e.g. engagement level) that have to be estimated separately. However, the tabular approaches like Q-learning (used in our recent work [89]), are unsuitable for big state spaces [2], like raw sensory data. That is why in this work, we are planning to use deep learning (DL) for creating a visuomotor behaviour model. In the next section, we will introduce DL techniques used in the context of decision-making for social robotics.

2 Problem Statement

The problem that is going to be approached during the project is to develop a personalised behaviour model for the social robot which will increase its autonomy during RAT for children with ASD. This should decrease the workload of the therapist that will not have to control the robot remotely. The model should be able to perform both social behaviour and game difficulty personalisation, as the children should increase their skills and stay engaged during the intervention.

As we want to avoid an error provided by an additional model, such as engagement estimator (calculating high-level features for the behaviour model), we want to develop a decision-making algorithm that operates on the raw sensory data. As the conventional tabular approaches are unsuitable for that purpose [2], we will focus on the DL techniques.

2.1 Deficits to Be Solved

First of all, based on our literature search, in HRI there are no DL based solutions for both social behaviour and game difficulty personalisation. That means that existing approaches for social behaviour adaptation are not designed for conducting games with a user. Other deficits of the social behaviour personalisation are related to the used models. First of all, the DL neural networks require a big amount of data in order to converge, that is why a big number of interactions with the user is needed. Moreover, the training is computationally heavy, that is why it can not be performed continuously during a human-robot interaction [61–63, 70]. The developed models for social behaviour adaptation in many cases were trained and evaluated on the previously collected datasets [16, 32, 69] or in the simulation [6, 7, 17]. Moreover, the data is often collected in the laboratory environment (instead of the real one), which might not prepare the model for the real-life interaction [16, 17, 69]. Additionally, the performance of the known solutions is not acceptable for the clinical intervention, due to the high possibility of making a mistake by the robot [16, 17, 32, 70]. This might be caused by the insufficient number of interactions with the user (not enough data) or the fact that the collected data is imbalanced (some actions are performed more often than others) [32]. Finally, many of the known solutions have a small action space which significantly limits the capabilities of the robot [6, 7, 16, 17, 61–63, 69, 70].

Secondly, up to our knowledge, in HRI there are no DL based solutions for game difficulty personalisation. There are only works describing applications of DRL algorithms as virtual [30, 55, 87] or robotic [20, 21] opponents for different kinds of games.

In this project we plan to face the following challenges:

- enabling the robot to personalise the game difficulty and social behaviour, by feeding to the DL network not only sensory data but also an additional information with an extra input vector (similarly to [7, 32, 63]) containing the game information,
- maintenance of an adequately big action space for conducting a therapy intervention (similar to [89, 90, 95]),

- making the model less erroneous, by pretraining the model on the manually collected dataset and applying learning from guidance approach, similarly to [82].

2.2 Proposed Approach

Our proposed approach would be an adpatation of the CNN used in [69, 70] or of SocialDQN [7]. In the second case, instead of an extra input (additional to the input with 8 grayscale images) in the form of one-hot vector encoding for the social signals, we suggest an extra input that would encode the current game state, as depicted in Fig. 1. This solution would provide the network with a necessary information for giving an engaging feedback to the user (based on the image), as well as adapting the difficulty of the game (based on the game signals). The game signals vector would contain the information about the last chosen difficulty level and if the user succeeded solving the task.

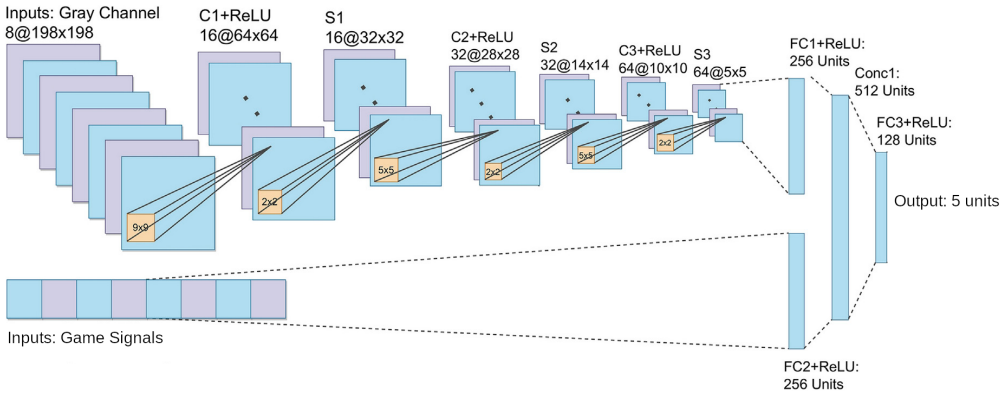

Figure 1: Adapted architecture of SocialDQN [7]

The action space will consist of 5 actions, similarly as in [89, 90, 95], namely, setting one of the three difficulty levels and providing encouraging or challenging feedback. Regarding the reward for the model, we suggest the similar approach as in [82], which is learning from guidance. We make an assumption that during the therapy for children with ASD, there is always a therapist that can accept or overwrite the actions of the robot (before being executed). We thus base the reward on the information from the therapist, who is a supervisor of the system in

that case. The feedback from the therapist will not only influence the reward, but also the action executed in the end. We believe, that similar changes (introducing an extra input vector with game signals, increasing the number of possible actions, retraining based on the feedback from the supervisor) can be done to the simpler model, which is a CNN presented in [69, 70].

2.3 Evaluation

We will evaluate the proposed approach on the manually collected data during the experiments conducted in the laboratory environment as in [89]. In order to increase the amount of training data we will apply augmentation techniques (e.g. mirroring the image in the vertical axis as mentioned in [68]).

3 Related Work[2]

As we discussed in [91], there are four personalisation approaches that are provided in the context of Human-Robot Interaction (HRI), namely: social behaviour, game difficulty, affective, and user preferences (e.g. proxemics) personalisation. However, the main focus should be put on two first personalisation techniques. The DL has been explored in the field of the social behaviour personalisation and game difficulty personalisation. The latter, however, refers more to the agent not as a tutor but as an opponent in the game.

3.1 Deep Learning Personalisation Approaches

3.1.1 Social Behaviour Personalisation

Qureshi et al [61] made a step towards making robots more interactive while coexisting with humans. This is done by enabling the robot to continuously learn social interaction skills. For that purpose, the deep reinforcement learning

[2]Parts of this chapter have been published in [91]

(DRL) method was used, which is called Multimodal Deep Q-Network (MDQN)[3][4], which is based on DQN [55]. This is a major contribution, as the developed DRL interaction model can conduct a human-robot interaction in an uncontrolled and real environment, which is not the case for the previous works. There are two inputs for MDQN, namely grayscale and depth video frames. Based on these two streams the model learns when to perform one of the four actions, which are: *wait*, *look towards human*, *wave hand*, *handshake*. The reward function is designed such that the robot's main goal is to perform a successful handshake. Here the data generation and training phases are separated, which means that all the robot's experiences are saved in the replay memory and used for training only during the resting period.

In [62] the MDQN model was augmented by adding a recurrent attention model [87] in order to enable the network to focus on certain fragments of the input data. This approach was proven to increase people's willingness to interact with a robot (through handshaking). However, the new model requires more training data in order to reach performance of the neural network without attention. There are three deficits of the aforementioned approaches which are related to deploying the robot in the real environment. Firstly, the action space consists only from four actions, which might be insufficient for reacting to human behaviour in real life. Secondly, the designed reward function is task-specific and may be difficult to adjust to different applications (here handshaking is rewarded explicitly); it should be mentioned as well that external rewards are scarce and might make learning relatively slow. Moreover, as training the neural network is computationally heavy, it has to be performed during the robot resting period, as otherwise it could cause delays during the interaction period. This is a significant disadvantage, which implies that the policy is not updated continuously during an interaction, which may cause a user to be disengaged.

In the follow-up work [63], the authors faced the second deficit and proposed an intrinsically motivated DRL. The internal reward, used for training the Deep Q-network (Fig. 2), is calculated with the use of an action-conditional prediction

[3]The original implementation is available under `https://github.com/ahq1993/Multimodal-Deep-Q-Network-for-Social-Human-Robot-Interaction`

[4]The python implementation (used in [6]) is available under `https://github.com/JPedroRBelo/pyMDQN`

network (Pnet) as an error between Pnet's event prediction and detected event occurrence. The aforementioned events are an observed behaviour of the interacting partner, such as handshake, eye contact and smile. It is moreover shown that intrinsically motivated DRL leads to more human-like behaviour than DRL based on an external reward. However, during the experiments it was noticed that the robot was repeating a handshake action even after already successful handshake. This behaviour may be unusual from the perspective of the normal person and should be avoided (e.g. with memory and by preventing the robot from being oblivious).

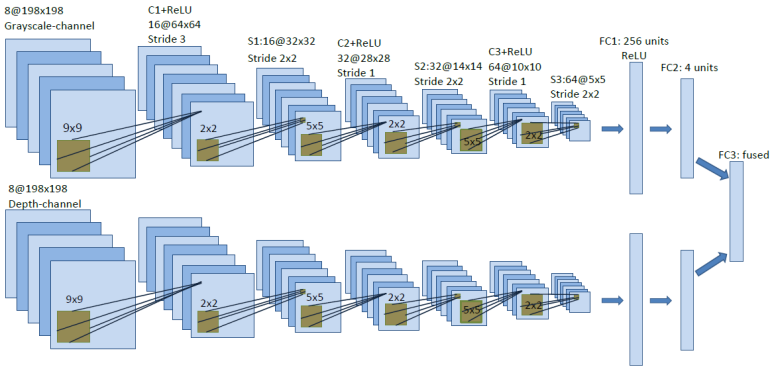

Figure 2: Architecture of the Deep Q-Network [63]

Belo et al [7] also made an approach to make robots more socially acceptable and natural while interacting with humans. For that purpose the Social Robotics Deep Q-Network (SocialDQN) was developed, which is based on the MDQN architecture. However, in SocialDQN the input for the depth information was substituted with the input indicating the emotional state of the user with Ekman's six basic emotions [26]: *fear*, *happiness*, *disgust*, *anger*, *sadness*, *surprise*. The authors conducted tests in the developed simulator [6] where the 13 human referees had to assess if the robot performed correct actions in each of the considered scenarios. In the end three models were compared, namely a random policy action selector, SocialDQN with and without social signals as an extra input. It was shown that SocialDQN, considering social signals, outperforms the two other predictors. It reached an accuracy of 89.50%. Unfortunately, no comparative study was performed with an original MDQN. Moreover, the evaluation was performed only in the simulation

and not during a real-life interaction with a real human. Another deficit is that the action space of SocialDQN is relatively small as it consists of only four actions (similarly to MDQN).

Clark-Turner et al [16] aim at developing a framework enabling the robot to deliver a behavioural intervention (BI). The idea behind BI is to teach children with certain disorders to perform new behaviours. The authors used for that purpose deep recurrent Q-network (DRQN) which is trained from human demonstrations. This is a major contribution as no previous work used DRQN to learn human policy for choosing discrete actions based on the demonstration data. There are two inputs for the developed model, namely: RGB image, point cloud and an audio spectrogram. The set of possible actions consists of: *command*, *prompt* (in both cases the robot waves and greets the user, with an exception that the latter provides also a prompt saying what is expected from the participant), *reward* (giving a positive appraisal to the user followed by ending the session by saying goodbye), *abort* (ending the session when the user's response is negative consequently). The reward for the robot is positive whenever the actions are performed correctly, otherwise no reward is provided. Unfortunately, no real-life tests were conducted as the proposed algorithm was evaluated only with the collected demonstration data, where the robot was tele-operated. Moreover, the obtained results show a relatively small accuracy of the model (43.2%-83.3%) which may be insufficient for deploying it for the real behavioural intervention.

Both of the aforementioned deficits were faced in the next work of the same authors [17]. First of all, the approach to increase a model's accuracy with transfer learning was made. It means that the network extracting features from the RGB data was a IRNV2 network pre-trained on the ImageNet. Additionally, the used modalities were changed, namely a point-cloud was replaced with an optical flow image. Unfortunately, no comparative evaluation between an old and new approach was performed, so the improvement of the introduced changes can not be directly shown. However, the authors faced the second deficit and performed not only evaluation on the collected dataset but also a real-life tests. The reported accuracy in both cases was not exceeding 80%. Moreover, the authors limited the number of actions to three, namely: *PMT*, *REW*, *END* (which correspond to *prompt*, *reward* and *abort* respectively).

In [69] the goal of the authors is to develop the robot that would accompany the elderly. For that purpose the authors developed a multimodal system that can make a decision on how and with whom it should start an interaction. There are two contributions of this work, namely, collection of the dataset focusing on initiation of the interaction and training a convolutional neural network (CNN) on it. The data was collected in two sessions, during the first one single users were interacting with the robot and during the second one two participants were acting in front of the robot. The aforementioned CNN was trained and evaluated on the collected dataset. The input of the network consists of the grayscale images and the output of three actions: *waiting*, *calling for attention* (waving and introducing), *starting the interaction* (asking a direct question). The reward was designed to be positive or 0 in case of the correct use of the action and negative otherwise. The deficit of that work is that the trained CNN was evaluated only on the dataset (not in the real-life interaction) and the limited action space (only three actions). However, the reported accuracy is promising as it reached 93%.

In the next work [70], Romeo et al faced one deficit of [69] and evaluated the model during a real-life tests that lasted for four days. To improve the models capabilities, at the end of each day, the network was fully retrained (including the new data). According to the reported results, the accuracy of the network, obtained on the updated dataset was still 93% at the end of the last experimentation day. However, the accuracy of the CNN, obtained during an actual interaction with the study participants and calculated as number of correctly chosen actions accordingly to the participants opinions, was 44%, 58%, 56% and 59% (each one obtained on each single day). The additional details of the conducted experiments and the model can be found in [68].

Hijaz et al directly approached the problem of increasing autonomy of the robot used during the therapies for individuals with ASD [32]. The authors propose a system for learning the therapist's verbal behaviour during the intervention delivery through Learning from Demonstration (LfD). There are two contributions of the aforementioned work. Firstly, in the other works the collected data used for training the model is collected in a simulation [6, 7, 17] or laboratory environment [16, 17, 69] which usually simplifies real-world interaction. In this work however, the data was collected in a clinical setting, during a real intervention, where the robot

was tele-operated by a therapist. Secondly, in the previous works the actions delivered by the robot were pre-scripted which is the cause of the learned behaviour being unnatural in the end. In [32] the actions were taken directly from the demonstrations given by therapists (as the robot was tele-operated during the intervention). The input to the used deep neural network (DNN) model was a child audio spectrogram and last therapist behaviour. The action space is relatively big in comparison to other works [6, 7, 16, 17, 61–63, 69, 70] as it consists of 9 actions, namely: saying that the robot is in one of six emotional states (angry, surprised, tired, scared, happy, sad), discriminative stimulus (asking the child how the robot is feeling followed by emotion presentation using body language), social praise (in case the child gives a correct answer during a game) and random actions (used to maintain child's engagement). The main deficit of the presented work is a very low accuracy (43.48%), which might be due to insufficient amount of collected data which was additionally imbalanced. Moreover, the system was not tested in the real-life scenario, but only on the collected dataset.

3.1.2 Game Difficulty Personalisation

When it comes to game difficulty personalisation, the DL algorithms were used so far as opponent players in the board/computer games. In [85] the authors managed to develop the algorithm (based on deep neural network) that was able to defeat a professional player in the game of Go. The DQN was proved to achieve a performance of professional games testers [55]. In [30] the authors added Long Short-Term Memory (LSTM) [33] layer in order to enable DQN to remember distant events. In [87], the authors managed to improve the performance of DQN by adding an attention mechanism, which also increased an interpretability of the network decisions (ability to determine on which regions of the image the network is focusing).

Finally, in the robotics field, in [20], the author designed another variant of DQN which was further deployed on the conversational robot playing the *Tic-tac-toe* game. The main aim of the work was to develop a humanoid robot that would play games with people having a brain illness, and thus slow down their decline. The deployed model is an improvement to the previous work as it provides not

only game difficulty but also social behaviour personalisation. The robot is able to play the game with the user, as well as perform certain dialogue acts. In total the action space is relatively big as it provides 18-34 actions (depending on the size of *Tic-tac-toe* grid), consisting of dialogue acts (speech and arm movements) and games moves. The state space consist of the features describing the game moves and words that occurred during an interaction. The authors evaluated the system with the user simulation (with semi-random behaviour). The results indicate that the system was able to conduct successful interactions and the proposed variant of DQN obtains higher winning rates than the original DQN.

The deficit of [20] was faced in [21] where the authors performed a real world evaluation. The proposed variant of DQN (in [21] extended to two variants of *Tic-tac-toe* game) was tested with 130 study participants for four nonconsecutive days in-the-wild (outside of the laboratory environment). The authors reported that the users were impressed with the robot, however no quantitative nor qualitative measures were provided.

However, all of the aforementioned works describe DRL agent as an opponent playing a game against a human user. No solutions for DRL just choosing the game difficulty for the user were found. Akalin et al [3] proposed an initial idea for a DRL for adjusting the difficulty level of the game during a social human-robot interaction with elderly people. According to the authors, the input would consist of three variables, namely, user's valence and engagement, as well as state of the game (the last difficulty level and information if the game is stopped or paused). The action space would consist of five actions: increasing, decreasing, not changing a difficulty level or pausing or stopping the game. The reward would be calculated based on valence and engagement of the user. Unfortunately, [3] is just a proposal of the solution without any implementation details nor results.

3.2 Available Datasets

For training a DL agents a huge amount of data is required, which explains a necessity for a dataset. We looked for publicly available datasets that can be used for training DL decision-making agents. Unfortunately, all of them can only be applied for testing a feasibility of social behaviour personalisation models. Because

the datasets contain visual data we categorised them into two groups, depending on the perspective of the camera from which the video/pictures were recorded. The following datasets contain data recorded from the first-person perspective (the human partner/s is/are directly interacting with the robot):

- AIR-Act2Act [45]: 100 subjects, 10 actions, 5000 samples, 3 modalities (depth, skeleton, RGB)
- NTU RGB+D [48, 84]: 106 subjects, 26 actions, 8276 samples, 3 modalities (depth, skeleton, RGB).
- [77]: 8 subjects, 7 actions, 684 samples, 1 modality (RGB)
- [76]: 8 subjects, 9 actions, 180 samples, 2 modalities (RGB, depth)
- AUTH UAV Gesture [58] (contains parts of datasets presented in [60, 84]): 8 subjects (newly captured), 6 actions, 4930 samples, 1 modality (RGB)

Most of the aforementioned datasets are designed for activity recognition [48, 77, 84], activity prediction [76] and gesture recognition [58]. Only the AIR-Act2Act [45] dataset was designed with the purpose of teaching the robot social skills, namely the scenario, where the study participant is performing an action, is labeled with the reaction type that the robot should undertake. Additionally, this dataset contains the robot behaviours (body gestures) that should be performed as a reaction. None of the aforementioned datasets can be used for training the model that would be applicable for ASD therapy. However, they can be used for checking a feasibility of the DL architecture (e.g. MDQN) that can be further trained in the simulation or on the manually collected dataset. Unfortunately, only certain datasets are easily available online, namely: AIR-Act2Act, AUTH UAV Gesture, NTU RGB+D.

The requirement for using a certain dataset are recordings from the first-person perspective (the person participating in the interaction). That is why the datasets containing recordings with only the perspective of the third person (two people are interacting with each other and the robot is only an observer) [34, 75, 97, 100] are unsuitable for our application.

(a) (b) (c) (d)

Figure 3: Frames (preprocessed) from one of the videos from AIR-Act2Act dataset [45] (the video depicts a scenario where the robot is supposed to bow to the elderly person who enters into the apartment)

3.3 Tools for Data Collection

As mentioned before, the provided datasets are not suitable for training the model that would be applicable for ASD therapy. That is why during the project it will be required to collect the dataset manually or in the simulation. In both cases the data will not be collected from the individuals with authism, however the scenarios imitating an robot-conducted intervention can be replicated with people without autism (students and university staff) or with simulated users. For both cases there are available tools that can be reused in the proposed project.

First of all, the autism intervention scenarios can be simulated with the Simulator for Deep Reinforcement Learning and Social Robotics (SimDRLSR) [6].[5] Here, the actions of the simulated humans are defined with scripts. Additionally, the robot's actions can be evaluated (whether they are socially acceptable) by human referees with a provided validation tool [7][6]. Although a simulation is a cheap and fast method for data collection for training a DL agent, the important features for action selection during an autism intervention are facial expressions, eye gaze and head orientation of the users [89], which might not be possible to reflect in the simulation.

Another possibility is to collect the data manually. This requires a suitable software that would enable a researcher to control the robot, as well as appropriately

[5] `https://github.com/JPedroRBelo/simDRLSR`
[6] `https://github.com/JPedroRBelo/validation_tool_socialdqn`

organise and store the collected data. There are publicly available packages that can be used as a guideline for creating such a software [11, 12, 17].[7,8]

4 Project Plan

4.1 Work Packages

During the master thesis the following working packages will be delivered:

WP1 **Literature search on the DL adaptation techniques used in HRI**
Objectives:

- Search the DL based social behaviour learning approaches commonly used in HRI.
- Search the DL based game difficulty adaptation techniques commonly used in HRI.

WP2 **Selection of the best performing DL architecture for a social behaviour learning**
Objectives:

- Select available DL based social behaviour learning architectures [7, 61, 62, 69, 70].
- Select suitable datasets [45, 48, 58, 84].
- Compare the previously selected DL architectures based on one of the previously selected datasets and choose the best performing model.

WP3 **Adaptation of the social behaviour DL model for the sequence-learning game use case**
Objectives:

- Adaptat the chosen DL architecture, such that it can be used for both social behaviour learning and game difficulty adaptation (the proposed approach is described in subsection 2.2).

[7] `https://github.com/AssistiveRoboticsUNH/deep_reinforcement_abstract_lfd`
[8] `https://github.com/AssistiveRoboticsUNH/TR-LfD`

- Perform initial training and evaluation of the adapted architecture on the dataset collected for our previous works [89, 90].

WP4 **Evaluation of the learning framework in the real-world environment**
Objectives:

- Prepare the experimentation setup and gather volunteers for the experiments.
- Conduct an evaluation with the proposed DL model (deployed on the QTrobot or NAO) similarly to [61, 68].

WP5 **Project report**
Objectives:

- Merge the results from the previous work packages.
- Write the final report.

4.2 Milestones

M1 **DL behaviour model selection**
At this milestone, the literature search in the field of DL in social behaviour learning and game difficulty adaptation should be done. Based on the evaluation of the selected social behaviour models the best performing one should be found. The respective results of literature search and models' evaluation should be summarized in the form of reports.

M2 **DL behaviour model adaptation and experimental setup**
At this milestone, the most adequate social behaviour model should be adapted to the sequence-learning game and initially evaluated with a dataset collected for our previous works. If the aforementioned dataset will not be sufficient (e.g. too small amount of samples) the model will be trained and evaluated only during the real-life evaluation, for which the experimentation setup should be prepared. At this milestone, the implementation of the required DL algorithm as well as learning/evaluation pipeline should be

ready. Additionally, the software for real-life evaluation should be prepared adequately.

M3 **Real-life evaluation and report submission**
At this milestone, the results of the real-life evaluation of the DL behaviour model should have been obtained. The final master thesis report, containing all the results from the previous milestones should be submitted.

4.3 Tasks and Deliverables

Tasks belonging to each working package as well as related deliverables are presented in the Table 1.

Working Package / Task	Description	Deliverable
WP1	**Literature search on the DL adaptation techniques used in HRI**	
T1.1	Research on DL based social behaviour learning concepts used in HRI	D1.1: Report with the overview of the commonly used methods
T1.2	Research on DL based game difficulty adaptation concepts commonly used in HRI	D1.2: Report with the overview of the commonly used methods
WP2	**Selection of the best performing DL architecture for a social behaviour learning**	
T2.1	Selection of the available DL based social behaviour personalisation architectures	D2.1: Report with the description of the selected architectures and dataset
T2.2	Selection of the suitable datasets	
T2.3	Comparative evaluation of the previously selected DL architectures based on one of the previously selected datasets and choice of the best performing model	D2.2: Implementation of the models and report with the evaluation results
WP3	**Adaptation of the social behaviour DL model for the sequence-learning game**	

T3.1	Adaptation of the chosen DL architecture, such that it can be used for both social behaviour learning and game difficulty adaptation	D3.1: Implemented software
T3.2	Initial training and evaluation of the adapted architecture on the dataset collected for our previous works	D3.1: Report with the evaluation results
WP4	**Evaluation of the learning framework in the real-world environment**	
T4.1	Preparation of the experimentation setup	D4.1: Implemented software
T4.2	Evaluation with the proposed DL model (deployed on the QTrobot or NAO)	D4.2: Evaluation report
WP5	**Project report**	
T5.1	Merge of the reports from the previous deliverables	D5.1: Draft of the master thesis report
T5.2	Detailed description of the real-world experimentation	
T5.3	Creation of the proper visualisations of the evaluation data	
T5.4	Iterative corrections with the supervisors	D5.2: Final master thesis report

Table 1: Tasks and deliverables (WP stands for working package, T for task, D for deliverable)

4.4 Project Schedule

The project schedule is illustrated as a Gantt chart (Figure 4). It includes the tasks as well as the milestones mentioned in the previous subsections.

References

[1] DE-ENIGMA Playfully Empowering Autistic Children. `https://de-enigma.eu/background-of-the-project/`. Accessed: 2022-02-05.

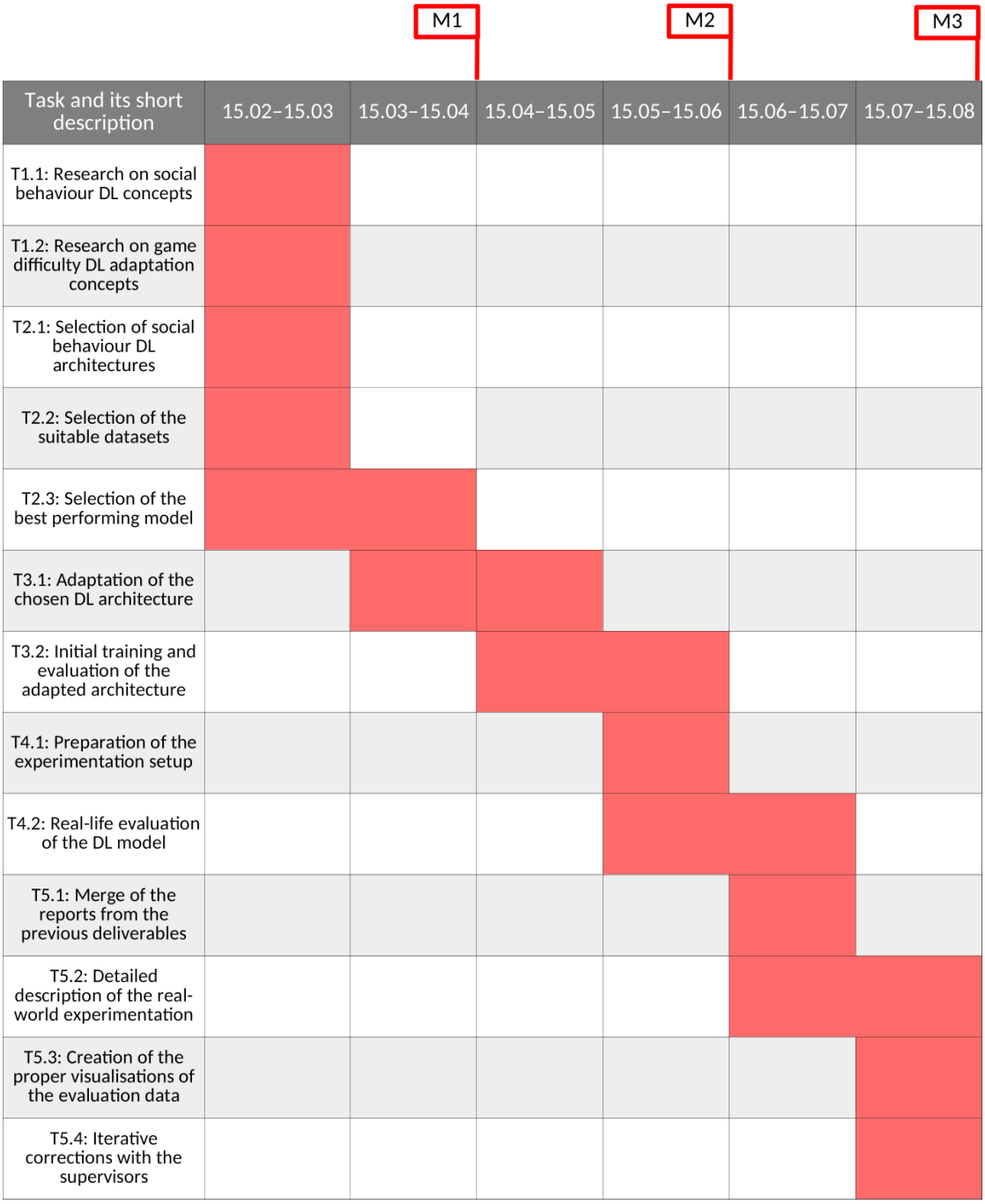

Figure 4: Project schedule

[2] Neziha Akalin and Amy Loutfi. Reinforcement learning approaches in social robotics. *Sensors*, 21(4):1292, 2021.

[3] Neziha Akalin, Andrey Kiselev, Annica Kristoffersson, and Amy Loutfi. Enhancing social human-robot interaction with deep reinforcement learning. In *FAIM/ISCA Workshop on Artificial Intelligence for Multimodal Human Robot Interaction*. ISCA, jul 2018. doi: 10.21437/ai-mhri.2018-12.

[4] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency. Openface 2.0: Facial Behavior Analysis Toolkit. In *13th IEEE Int. Conf. Automatic Face & Gesture recognition*, pages 59–66, 2018.

[5] Paul Baxter, Emily Ashurst, Robin Read, James Kennedy, and Tony Belpaeme. Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS one*, 12(5):e0178126, 2017.

[6] Jose Pedro R. Belo and Roseli A. F. Romero. A social human-robot interaction simulator for reinforcement learning systems. In *2021 20th International Conference on Advanced Robotics (ICAR)*. IEEE, dec 2021. doi: 10.1109/icar53236.2021.9659388.

[7] José Pedro R. Belo, Helio Azevedo, Josué J. G. Ramos, and Roseli A. F. Romero. Deep q-network for social robotics using emotional social signals. *Frontiers in Robotics and AI*, 9, sep 2022. doi: 10.3389/frobt.2022.880547. URL `https://github.com/JPedroRBelo/SocialDQN`.

[8] Hoang-Long Cao et al. A survey on behavior control architectures for social robots in healthcare interventions. *Int. Journal of Humanoid Robotics*, 14 (04):1750021, 2017.

[9] Hoang-Long Cao et al. A personalized and platform-independent behavior control system for social robots in therapy: development and applications. *IEEE Trans. Cognitive and Developmental Systems*, 11(3):334–346, 2018.

[10] Hoang-Long Cao et al. Robot-enhanced therapy: Development and validation of supervised autonomous robotic system for autism spectrum disorders therapy. *IEEE Robotics & Automation Magazine*, 26(2):49–58, 2019.

[11] Estuardo Carpio, Madison Clark-Turner, and Momotaz Begum. Learning sequential human-robot interaction tasks from demonstrations: The role of

temporal reasoning. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–8, 2019. doi: 10.1109/RO-MAN46459.2019.8956346.

[12] Estuardo Rene Carpio Mazariegos. Learning temporal dynamics of human-robot interactions from demonstrations. Master's thesis, 2018.

[13] Jeanie Chan and Goldie Nejat. Social intelligence for a robot engaging people in cognitive training activities. *Int. Journal of Advanced Robotic Systems*, 9 (4):113, 2012.

[14] Caitlyn Clabaugh and Maja Matarić. Escaping oz: Autonomy in socially assistive robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:33–61, 2019.

[15] Caitlyn Clabaugh et al. Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. *Frontiers in Robotics and AI*, page 110, 2019.

[16] Madison Clark-Turner and Momotaz Begum. Deep recurrent q-learning of behavioral intervention delivery by a robot from demonstration data. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, aug 2017. doi: 10.1109/roman.2017.8172429.

[17] Madison Clark-Turner and Momotaz Begum. Deep reinforcement learning of abstract reasoning from demonstrations. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 160–168, 2018.

[18] Andreia P. Costa, Georges Steffgen, Francisco J. Rodríguez Lera, Aida Nazarikhorram, and Pouyan Ziafati. Socially assistive robots for teaching emotional abilities to children with autism spectrum disorder. In *3rd Workshop on Child-Robot Interaction at HRI*, 2017.

[19] Andreia P. Costa, Georges Steffgen, Francisco J. Rodríguez Lera, Aida Nazarikhorram, and Pouyan Ziafati. Socially assistive robots for teaching

emotional abilities to children with autism spectrum disorder. In *3rd Workshop on Child-Robot Interaction at HRI*, 2017.

[20] Heriberto Cuayahuitl. Deep reinforcement learning for conversational robots playing games. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. IEEE, nov 2017. doi: 10.1109/humanoids.2017.8246959.

[21] Heriberto Cuayáhuitl. A data-efficient deep learning approach for deployable multimodal social robots. *Neurocomputing*, 396:587–598, jul 2020. doi: 10.1016/j.neucom.2018.09.104.

[22] Daniel O David, Cristina A Costescu, Silviu Matu, Aurora Szentagotai, and Anca Dobrean. Developing joint attention for children with autism in robot-enhanced therapy. *Int. Journal of Social Robotics*, 10(5):595–605, 2018.

[23] Francesco Del Duchetto and Marc Hanheide. Learning on the job: Long-term behavioural adaptation in human-robot interactions. *IEEE Robotics and Automation Letters*, 2022.

[24] Francesco Del Duchetto, Paul Baxter, and Marc Hanheide. Are you still with me? Continuous engagement assessment from a robot's point of view. *Frontiers in Robotics and AI*, 7:116, 2020.

[25] Thomas G Dietterich. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, 13:227–303, 2000.

[26] Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.

[27] Pablo G. Esteban et al. 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, 2017.

[28] Yongli Feng, Qingxuan Jia, and Wei Wei. A control architecture of robot-assisted intervention for children with autism spectrum disorders. *Journal of Robotics*, 2018, 2018.

[29] Goren Gordon et al. Affective personalization of a social robot tutor for children's second language skills. In *Proc. AAAI Conf. Artificial Intelligence*, volume 30, 2016.

[30] Matthew Hausknecht and Peter Stone. Deep recurrent q-learning for partially observable mdps. In *2015 aaai fall symposium series*, 2015.

[31] Jacaueline Hemminahaus and Stefan Kopp. Towards adaptive social behavior generation for assistive robots using reinforcement learning. In *2017 12th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI*, pages 332–340. IEEE, 2017.

[32] Ala'aldin Hijaz, Jessica Korneder, and Wing-Yue Geoffrey Louie. In-the-wild learning from demonstration for therapies for autism spectrum disorder. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 1224–1229, 2021. doi: 10.1109/RO-MAN50785.2021.9515439.

[33] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[34] Tao Hu, Xinyan Zhu, Wei Guo, and Kehua Su. Efficient interaction recognition through positive action representation. *Mathematical Problems in Engineering*, 2013:1–11, 2013. doi: 10.1155/2013/795360.

[35] Iolanda Iacono, Hagen Lehmann, Patrizia Marti, Ben Robins, and Kerstin Dautenhahn. Robots as social mediators for children with Autism - A preliminary analysis comparing two different robotic platforms. In *IEEE Int. Conf. Development and Learning (ICDL)*, volume 2, pages 1–6, 2011.

[36] Shomik Jain, Balasubramanian Thiagarajan, Zhonghao Shi, Caitlyn Clabaugh, and Maja J Matarić. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Science Robotics*, 5(39), 2020.

[37] Hifza Javed, Rachael Burns, Myounghoon Jeon, Ayanna M Howard, and Chung Hyuk Park. A robotic framework to facilitate sensory experiences

for children with autism spectrum disorder: A preliminary study. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(1):1–26, 2019.

[38] Abir B Karami, Karim Sehaba, and Benoit Encelle. Adaptive artificial companions learning from users' feedback. *Adaptive Behavior*, 24(2):69–86, 2016.

[39] Shofiyati Nur Karimah, Teruhiko Unoki, and Shinobu Hasegawa. Implementation of Long Short-Term Memory (LSTM) Models for Engagement Estimation in Online Learning. In *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*, pages 283–289. IEEE, 2021.

[40] Amanjot Kaur, Bishal Ghosh, Naman Deep Singh, and Abhinav Dhall. Domain adaptation based topic modeling techniques for engagement estimation in the wild. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–6. IEEE, 2019.

[41] Pooya Khorrami, Vuong Le, John C Hart, and Thomas S Huang. A system for monitoring the engagement of remote online students using eye gaze estimation. In *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2014.

[42] W Bradley Knox and Peter Stone. Interactively shaping agents via human reinforcement: The TAMER framework. In *Proceedings of the fifth international conference on Knowledge capture*, pages 9–16, 2009.

[43] W Bradley Knox and Peter Stone. Combining manual feedback with subsequent MDP reward signals for reinforcement learning. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 5–12. Citeseer, 2010.

[44] W Bradley Knox and Peter Stone. Reinforcement learning from simultaneous human and MDP reward. In *AAMAS*, pages 475–482, 2012.

[45] Woo-Ri Ko, Minsu Jang, Jaeyeon Lee, and Jaehong Kim. AIR-act2act: Human–human interaction dataset for teaching non-verbal social behaviors to robots. *The International Journal of Robotics Research*, 40(4-5):691–697, jan

2021. doi: 10.1177/0278364921990671. URL `https://github.com/ai4r/AIR-Act2Act`.

[46] Alyssa Kubota and Laurel D Riek. Methods for robot behavior adaptation for cognitive neurorehabilitation. *Annual Review of Control, Robotics, and Autonomous Systems*, 5, 2021.

[47] Daniel Leyzberg, Samuel Spaulding, and Brian Scassellati. Personalizing robot tutors to individuals' learning differences. In *2014 9th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, pages 423–430. IEEE, 2014.

[48] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C. Kot. NTU RGB+D 120: A large-scale benchmark for 3d human activity understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2684–2701, oct 2020. doi: 10.1109/tpami.2019.2916873.

[49] Sudhir S Mane and Anil R Surve. Engagement detection using video-based estimation of head movement. In *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pages 1745–1749. IEEE, 2018.

[50] Elisabeta Marinoiu, Mihai Zanfir, Vlad Olaru, and Cristian Sminchisescu. 3d human sensing, action and emotion recognition in robot assisted therapy of children with autism. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2158–2167, 2018.

[51] Gonçalo S Martins, Luís Santos, and Jorge Dias. Bum: Bayesian user model for distributed social robots. In *2017 26th IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1279–1284. IEEE, 2017.

[52] Gonçalo S Martins, Luís Santos, and Jorge Dias. BUM: Bayesian User Model for Distributed Learning of User Characteristics From Heterogeneous Information. *IEEE Transactions on Cognitive and Developmental Systems*, 11(3):425–434, 2018.

[53] Gonçalo S Martins, Hend Al Tair, Luís Santos, and Jorge Dias. αPOMDP: POMDP-based user-adaptive decision-making for social robots. *Pattern Recognition Letters*, 118:94–103, 2019.

[54] Gonçalo S Martins, Luís Santos, and Jorge Dias. User-adaptive interaction in social robots: A survey focusing on non-physical interaction. *Int. Journal of Social Robotics*, 11(1):185–205, 2019.

[55] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[56] Olivia Nocentini et al. A survey of behavioral models for social robots. *Robotics*, 8(3):54, 2019.

[57] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1990.

[58] Fotini Patrona, Ioannis Mademlis, and Ioannis Pitas. An overview of hand gesture languages for autonomous uav handling. *2021 Aerial Robotic Systems Physically Interacting with the Environment (AIRPHARO)*, pages 1–7, 2021.

[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830, 2011.

[60] Asanka G. Perera, Yee Wei Law, and Javaan Chahl. Uav-gesture: A dataset for uav control and gesture recognition. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[61] Ahmed Hussain Qureshi, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Robot gains social intelligence through multimodal deep reinforcement learning. In *2016 IEEE-RAS 16th International Conference on*

Humanoid Robots (Humanoids). IEEE, nov 2016. doi: 10.1109/humanoids.2016.7803357.

[62] Ahmed Hussain Qureshi, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 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, may 2017. doi: 10.1109/icra.2017.7989193.

[63] Ahmed Hussain Qureshi, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Intrinsically motivated reinforcement learning for human–robot interaction in the real-world. *Neural Networks*, 107:23–33, nov 2018. doi: 10.1016/j.neunet.2018.03.014.

[64] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[65] Hannes Ritschel, Tobias Baur, and Elisabeth André. Adapting a robot's linguistic style based on socially-aware reinforcement learning. In *2017 26th ieee international symposium on robot and human interactive communication (ro-man)*, pages 378–384. IEEE, 2017.

[66] Ben Robins, Kerstin Dautenhahn, and Janek Dubowski. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction Studies*, 7(3):479–512, 2006.

[67] Ben Robins, Kerstin Dautenhahn, Luke Wood, and Abolfazl Zaraki. Developing interaction scenarios with a humanoid robot to encourage visual perspective taking skills in children with autism–preliminary proof of concept tests. In *Int. Conf. on Social Robotics*, pages 147–155. Springer, 2017.

[68] Marta Romeo. *Human-robot Interaction and Deep Learning for Companionship in Elderly Care*. PhD thesis, University of Manchester, 2021. URL `https://www.research.manchester.ac.uk/portal/files/194692310/FULL_TEXT.PDF`.

[69] Marta Romeo, Angelo Cangelosi, and Ray Jones. Developing a deep learning agent for HRI: Dataset collection and training. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, aug 2018. doi: 10.1109/roman.2018.8525512.

[70] Marta Romeo, Daniel Hernández García, Ray Jones, and Angelo Cangelosi. Deploying a deep learning agent for HRI with potential "end-users" at multiple sheltered housing sites. In *Proceedings of the 7th International Conference on Human-Agent Interaction*. ACM, sep 2019. doi: 10.1145/3349537.3351886.

[71] Silvia Rossi, Francois Ferland, and Adriana Tapus. User profiling and behavioral adaptation for HRI: A survey. *Pattern Recognition Letters*, 99:3–12, 2017.

[72] Ognjen Rudovic, Jaeryoung Lee, Lea Mascarell-Maricic, Björn W Schuller, and Rosalind W Picard. Measuring engagement in robot-assisted autism therapy: a cross-cultural study. *Frontiers in Robotics and AI*, 4:36, 2017.

[73] Ognjen Rudovic, Jaeryoung Lee, Miles Dai, Björn Schuller, and Rosalind W Picard. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, 3(19), 2018.

[74] M S Ryoo and J K Aggarwal. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *2009 IEEE 12th International Conference on Computer Vision*. IEEE, sep 2009. doi: 10.1109/iccv.2009.5459361.

[75] M. S. Ryoo and J. K. Aggarwal. UT-Interaction Dataset, ICPR contest on Semantic Description of Human Activities (SDHA), 2010. URL `http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html#Data`.

[76] M. S. Ryoo, Thomas J. Fuchs, Lu Xia, J. K. Aggarwal, and Larry Matthies. Robot-centric activity prediction from first-person videos: What will they do to me? In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 295–302, 2015.

[77] Michael S. Ryoo and Larry Matthies. First-person activity recognition: What are they doing to me? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[78] Brian Scassellati et al. Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*, 3(21), 2018.

[79] Emmanuel Senft. *Teaching Robots Social Autonomy From In Situ Human Supervision*. PhD thesis, University of Plymouth, 2018.

[80] Emmanuel Senft, Paul Baxter, and Tony Belpaeme. Human-guided learning of social action selection for robot-assisted therapy. In *Machine Learning for Interactive Systems*, pages 15–20. PMLR, 2015.

[81] Emmanuel Senft, Paul Baxter, James Kennedy, and Tony Belpaeme. Sparc: Supervised progressively autonomous robot competencies. In *Int. Conf. Social Robotics ICSR*, pages 603–612. Springer, 2015.

[82] Emmanuel Senft, Paul Baxter, James Kennedy, Séverin Lemaignan, and Tony Belpaeme. Supervised autonomy for online learning in human-robot interaction. *Pattern Recognition Letters*, 99:77–86, 2017.

[83] Emmanuel Senft, Séverin Lemaignan, Paul E Baxter, Madeleine Bartlett, and Tony Belpaeme. Teaching robots social autonomy from in situ human guidance. *Science Robotics*, 4(35), 2019.

[84] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. URL `https://rose1.ntu.edu.sg/dataset/actionRecognition/`.

[85] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the

game of go with deep neural networks and tree search. *Nature*, 529(7587): 484–489, jan 2016. doi: 10.1038/nature16961.

[86] Satinder P Singh and Richard S Sutton. Reinforcement learning with replacing eligibility traces. *Machine learning*, 22(1):123–158, 1996.

[87] Ivan Sorokin, Alexey Seleznev, Mikhail Pavlov, Aleksandr Fedorov, and Anastasiia Ignateva. Deep attention recurrent q-network. *arXiv preprint arXiv:1512.01693*, 2015.

[88] Michał Stolarz, Alex Mitrevski, Mohammad Wasil, and Paul G. Plöger. Personalised Behaviour Model for Autism Therapy. In *ICRS Workshop on Social AI for Human-Robot Interaction of Human-Care Robots*, 2021.

[89] Michał Stolarz, Alex Mitrevski, Mohammad Wasil, and Paul G. Plöger. Learning-Based Personalisation of Robot Behaviour for Assistive Robotics. *IEEE Transactions on Robotics*, 2022.

[90] Michał Stolarz, Alex Mitrevski, Mohammad Wasil, and Paul G. Plöger. Personalised Robot Behaviour Modelling for Robot-Assisted Therapy in the Context of Autism Spectrum Disorder. In *RO-MAN Workshop on Behavior Adaptation and Learning for Assistive Robotics*, 2022.

[91] Michał Stolarz, Alex Mitrevski, Mohammad Wasil, and Paul G. Plöger. Personalized Behaviour Models: A Survey Focusing on Autism Therapy Applications. In *HRI Workshop on Lifelong Learning and Personalization in Long-Term Human-Robot Interaction (LEAP-HRI)*, 2022.

[92] Lisa Torrey and Matthew Taylor. Teaching on a budget: Agents advising agents in reinforcement learning. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 1053–1060, 2013.

[93] Konstantinos Tsiakas, Maher Abujelala, Alexandros Lioulemes, and Fillia Makedon. An intelligent interactive learning and adaptation framework for robot-based vocational training. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6. IEEE, 2016.

[94] Konstantinos Tsiakas, Maria Dagioglou, Vangelis Karkaletsis, and Fillia Makedon. Adaptive robot assisted therapy using interactive reinforcement learning. In *Int. Conf. on Social Robotics*, pages 11–21. Springer, 2016.

[95] Konstantinos Tsiakas, Maher Abujelala, and Fillia Makedon. Task engagement as personalization feedback for socially-assistive robots and cognitive training. *Technologies*, 6(2):49, 2018.

[96] Konstantinos Tsiakas, Maria Kyrarini, Vangelis Karkaletsis, Fillia Makedon, and Oliver Korn. A taxonomy in robot-assisted training: current trends, needs and challenges. *Technologies*, 6(4):119, 2018.

[97] Coert van Gemeren, Ronald Poppe, and Remco C. Veltkamp. Spatio-temporal detection of fine-grained dyadic human interactions. In *Human Behavior Understanding*, pages 116–133. Springer International Publishing, 2016. doi: 10.1007/978-3-319-46843-3_8.

[98] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8 (3-4):279–292, 1992.

[99] Katie Winkle, Severin Lemaignan, Praminda Caleb-Solly, Paul Bremner, Ailie Turton, and Ute Leonards. In-situ learning from a domain expert for real world socially assistive robot deployment. In *Proceedings of Robotics: Science and Systems*, volume 10, 2020.

[100] Kiwon Yun, Jean Honorio, Debaleena Chattopadhyay, Tamara L. Berg, and Dimitris Samaras. Two-person interaction detection using body-pose features and multiple instance learning. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, jun 2012. doi: 10.1109/cvprw.2012.6239234.