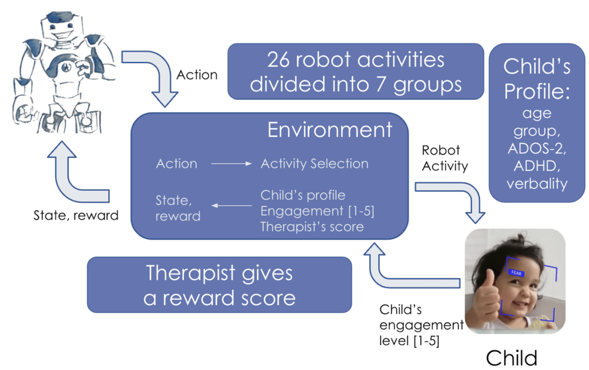
ROBT491 Graduation Project: Adapting to Learner's Cognitive Differences Using Reinforcement Learning

Ilyas Issa[1], Saparkhan Kassymbekov[2], Zholaman Kuangaliyev[3], Symbat Nurgazy[4]
[1]ilyas.issa@nu.edu.kz, [2]saparkhan.kassymbekov@nu.edu.kz, [3]zholaman.kuangaliyev@nu.edu.kz, [4]symbat.nurgazy@nu.edu.kz

***Abstract*—This report explores the benefits and challenges of creating adaptive robotic systems by integrating open-source software and reinforcement learning algorithms. The aim is to develop robotic systems that adapt to the cognitive differences of the child to increase engagement time and develop social skills using robot-assisted therapy. The Furhat and NAO robots are used as the platforms for receiving inputs and outputs, while a reinforcement learning agent selects the order of activities. The state space includes the pose landmarks from the Mediapipe and OpenPose libraries to recognize the engagement level and some static parameters such as age, autism level, verbality, and the co-diagnosis of Attention Deficit Hyperactivity Disorder. The actions space contains the act of changing the current activity. In the current stage, we have eight types of activities the robot chooses to interact with the child. Furhat robot is programmed to have a full pipeline for a conversational robot: automatic speech recognition, text-to-speech, machine translation, and OpenAI language model. The results of this project provide insight into the potential of using open-source software and reinforcement learning algorithms to create more advanced interactive robots and highlight the importance of continued research in the field of Human-Robot Interaction.**



***Index Terms*—reinforcement learning, child-robot interaction, robot-assisted autism therapy**

I. Introduction

Autism intervention should begin as early as possible after receiving a clinical diagnosis of Autism Spectrum Condition (ASC). There are a variety of treatment options available globally that focus on social learning, such as behavioural, developmental, educational, socio-relational, pharmacological and etc. [1]. However, autism is still not well understood by the general population in Kazakhstan, and there is often a lack of resources and services available for children with autism and their families [2]. Therefore, there is a pressing need to provide comprehensive and effective support for individuals with autism and their families. Currently, specialized centers for autism therapy are developing and expanding in Kazakhstan, which presents an opportunity to introduce and implement adaptive robotic systems for Robot-Assisted Autism Therapy (RAAT) in the region. [2]. There is no single method that can cure autism, and thus, the primary goal of adaptive robotic systems is to reduce autistic behavioural patterns, enabling individuals to live a more independent life and build social relationships. While some children make remarkable gains in improving social and communication skills, others make slow or limited progress [3]. Thus, a special focus on those children who do not progress is required.

Our primary focus rests on behavioral treatment which helps children to develop social skills for daily life. We bring this aspect to an adaptive system that focuses on developing social, behavioural, and emotional learning while interacting with robots in different social environments. This area is yet at the very beginning of development because most studies remain cross-sectional and lack generalization. Social robots are considered to engage children in interactive learning activities that supplement and augment those provided by therapists [4]. However, behavioural research requires large data on children to optimize their learning. It currently lacks diverse data from children with ASC due to the small sample size in experimental research [5]. Another related aspect is the extent to which current human-robot interaction (HRI) studies can ensure tailored support and that the shift from teleoperated to autonomous interactions brings challenges in computer-based perception and robot control [6].

The main objectives of the project can be divided into two categories: scientific objectives and real-world application objectives. The scientific objective is to investigate the potential of using reinforcement learning to develop socially adaptive robot applications for robot-assisted autism therapy. Reinforcement learning is a subgroup of machine learning in which an agent learns to make decisions by performing actions and receiving rewards or penalties in response to the task outcome [7]. The real-world application objective is to develop robot applications on Furhat and NAO robots and integrate/introduce them at the Kazakhstani center for children with autism.

II. Related Works

Autism therapy commonly offers highly structured and predictive interventions to accommodate learners' needs. Yet, some promising efforts towards adaptive and autonomous systems have been already taken in HRI. Past research shows a greater need for game difficulty personalization to fit each child's skill level and to prevent them from getting demotivated during intervention [8]. For instance, Scasselatti et al. [6] integrated personalization in RAAT by adjusting the difficulty levels of practice games, similar to the personalization of educational content in other smart systems. Over a period of one month, they deployed autonomous robots for a total of 127 hours over 279 sessions at the homes of children with ASC. As a result, they found that practising joint attention with the robot could help children to transfer those skills when interacting with other people.

In the work by Zini et. al. called "Adaptive Cognitive Training with Reinforcement Learning", authors suggest an approach for adaptive cognitive training that is based on reinforcement learning (RL) [9]. In order to maximize training efficacy, the method employs an RL algorithm to dynamically modify the complexity of training tasks based on the user's performance. The authors demonstrate that their RL-based approach leads to superior performance compared to a predetermined training regimen by evaluating it on a variety of cognitive training tasks. They also demonstrate how the strategy effectively accommodates individual variances, resulting in more customized training plans. Overall, the findings point to RL as a potentially effective method for increasing cognitive training.

Another work by Shawki and Badawi examines the use of RL to provide a personalized learning experience. The authors suggest a system that creates a personalized learning environment by adjusting to the user's learning style, motivation, and speed. The technology adapts the learning experience based on user input and reinforcement learning algorithms to understand individual preferences. Through simulations and trials with a sample of users, the authors show the usefulness of the suggested approach. The findings demonstrate that the system is capable of delivering a tailored learning experience that raises motivation, engagement, and performance [10].

The paper "Reinforcement Learning Approaches in Social Robotics" by Neziha Akalin and Amy Loutfi [11] provides a comprehensive review of the works on using the reinforcement learning algorithm in social robotics. The authors argue that the use of RL might improve the social skills and interactions of robots with humans by addressing the transparency issues through social cues. Also, the authors suggest the use of Iterative Reinforcement learning (IRL) and a complex reward system in order to deal with the sparse reward problem, when the algorithm does not receive a reward due to slow change in a person's cues.

Park et. al. proposes a model-free affective reinforcement learning approach to personalize an autonomous social robot companion for early literacy education. With this method, the robot may learn the preferences of the youngster it is working with and modify its behaviour appropriately. Reinforcement learning algorithms are used. By customizing its interactions with the child based on their emotional state, the goal is to increase the robot's efficacy as a teaching tool. The study's findings that the suggested strategy works well at increasing children's motivation and involvement with educational activity are presented in the article [12].

The work "Comparing the effectiveness of different reinforcement stimuli in a robotic therapy for children with asd" focuses on developing a reinforcement learning approach for optimizing the interaction between a custom-designed socially assistive robot and children with autism spectrum disorder [13]. The authors use the actor-critic algorithm and propose a reward function that considers both the child's behavior and the robot's actions. Simulation results demonstrate the effectiveness of the proposed approach in improving the interaction between the child and the robot. The authors argue that this reinforcement learning approach has the potential to provide a more personalized and effective therapy experience for children with autism spectrum disorders compared to traditional methods. This related work offers insight into the application of reinforcement learning in socially assistive robots for children with autism spectrum disorders and its potential benefits in terms of interaction and personalization.

In the study "Towards robot-assisted therapy for children with autism—the ontological knowledge models and reinforcement learning based algorithms," the authors present a reinforcement learning framework for a robot to be used in adaptive robot-assisted therapy [14]. The authors use the learning algorithm to create a system that predicts a learner's response to various therapy interventions and adjusts the therapy plan accordingly. They propose a state-action-reward model that takes into consideration the learner's current state and the impact of different therapeutic interventions. Simulation studies demonstrate the efficacy of the proposed model in adapting to different learners' needs and improving therapy outcomes. The authors conclude that reinforcement learning has a broad range of applications in robot-assisted therapy, offering a more personalized and effective approach compared to traditional, one-size-fits-all methods. This related work provides a valuable understanding of the use of reinforcement learning in robot-assisted therapy using a robotic arm, and the potential benefits it brings in terms of personalization and effectiveness compared to traditional methods.

"Designing a Socially Assistive Robot for Long-Term In-Home Use for Children with Autism Spectrum Disorders" paper describes the design and implementation of the socially assistive robot, intended for long-term in-home use with children with autism spectrum disorders [15]. The authors aim to create a robot that can engage children in various therapy activities and improve their social and communication skills. Robot uses a combination of machine learning and human input to adapt to the child's preferences and behavior over time. Results show the effectiveness of this approach in providing a more personalized therapy experience and improving the child's skills. This related work highlights the potential of socially assistive robots for therapy for children with autism spectrum disorders.

Affect recognition is another aspect of HRI contribution.

Creating deep learning models that can accurately predict a user's state during autism therapy is still a challenge. Some studies address this issue with the help of adaptive systems. For instance, Rudovic et al. [4] developed a personalized deep learning framework, the personalized perception of affect network (PPA-net), that can adapt robot perception of children's affective states and engagement to different cultures and individuals. This multi-modal collects data from unobtrusive sensors and consists of video recordings of facial expressions, body movements, audio data, and physiological data such as heart rate. The framework operates in three steps: sensing, perception and interaction. In their follow-up study [16], the data selection method was found to be effective for new children. This is because it can choose the appropriate data samples that allow the pre-trained engagement classifiers to quickly adapt to the new child. Personalizing the multi-modal model can significantly improve the accuracy of engagement assessment for each child, even when using only a small amount of data. The authors stress that data labelling is both time-consuming and expensive and that it can also be affected by human bias when dealing with multi-modal data. Next up, Clabaugh et al [17] proposed RAAT personalization as a hierarchical human-robot learning framework (hHRL) with five controllers, disclosure, promise, instruction, feedback, and inquiry, to personalize instruction challenge levels and robot feedback as per each user's learning patterns. They validated the framework with 17 children with ASD, aged 3–7 years old, over one month in their homes. The results show that the fully autonomous robot system could personalize its instruction and feedback to each child's needs over time, with improvements in targeted skills and post-intervention retention of content.

III. Reinforcement learning for RAAT

To optimize engagement and learning, the robots can use RL to track a child's learning patterns and use that information to modify the level of task difficulty. The adaptation might also benefit RL to observe the child's engagement in real-time and adapt its social behaviours. For instance, the robot could modify the activity or switch to another if the child works on a challenging and uninteresting task. RL can personalize the learning content to be effective and engaging. The proposed RL-based system adapts to each child's individual characteristics by choosing tasks that children would need to practice and improve targeted skills. The RL algorithm mediates between the robot and the child by collecting task performance data through the internal and external sensors of the system. Additionally, it communicates information to the robots, awaits the response, and then ultimately acts on it to increase user engagement.

A. Environment

We created our own custom environment based on the OpenAI Gym library.

- **State space**: The state space of the environment has two parts. The first part is the variables that are static during the full cycle of interaction with one child. These are age, the presence of ADOS2, ADHD ASD, the numerical data on whether the child is compliant or non-compliant, or whether verbal or nonverbal. They are predefined before the interaction. The second part includes variables of the state space that change during the interaction. We have tried multiple variants of such data. These variants include the binary engagement level, engagement level from 0 to 5, and 412 pose landmarks from OpenPose. We tried each, one by one, to be the state space and compared the results. So the size of the observation varies from 7 to 420.
- **Action space**: The action space is based on eight discrete actions. These actions are a change between the activity blocks of interaction such as "Dances", "Songs", "Touch Me", "Social Acts", "Storytelling", "Emotions", "Imitations" and "Hello and Bye". More detailed information about these activity blocks is written in Section IV-A. As not all the children had participated in all activities in our dataset, action space differs according to the session id and the child id of the current state.
- **Reward**: A reward for each step is calculated considering both static and dynamic parameters of the given state. The major dynamic parameter that we consider is the engagement level, both binary and from 1 to 5. If the engagement changes from 0 to 5, the value added to the reward changes from -5 to 0 respectively. The static parameters define a crucial aspect of the reward that changes unwanted actions. Depending on the values of those parameters, we define what values to add to the reward. For example, children with ADHD find it hard to participate in the "Storytelling" activity block, as it needs some level of concentration [18]. In that case, the reward is going to be largely negative for the "Storytelling" action. The reward values from the static and dynamic parameters are added together to give a final reward for the step. The target state for an agent is when the overall reward for the step is higher than zero.

The environment for training works in the following way. First, we reset the environment and randomly chose a state with the initial frame. Then during the training, in each step, we iterate through frames. For each step, our policy chooses an action, which has the maximum Q-value. The size of the Q-value is defined by the Deep Q-network. The action after being performed in the environment returns the next state, reward and the Boolean statement whether the agent has reached a target state or not.

B. Deep Q-Network Algorithm

We implement the Deep Q-Network (DQN) algorithm model to develop our system. The algorithm suits well to our project implementation because our environment has a relatively simple multidimensional state space and discrete action space. In other words, 8-420 plus 6 static parameters for state space and eight actions.

The algorithm deploys a neural network, made of linear layers, that extracts the feature vectors from the parameters of the state space and calculates the Q-value for each possible action in the action space. The action with the maximum Q-value is chosen by the agent as the action for a current state.

During training, the agent interacts with the environment and changes the Q-function until it finds the most optimal function by minimizing the difference between the target value derived from the Q-values of the next state and the actual Q-value. This is done by applying Mean Squared Error (MSE) method from Figure 1 with the Adam optimizer for training. The target value calculation is done by the formula from Figure 2.

$$L_i(\theta_i) = E_{(s,a,r,s')}\left[\left(y_i - Q\left(s,a;\theta_i\right)\right)^2\right]$$

Fig. 1: Loss function of the training algorithm [19]

$$y_i = r + \gamma \max_{a'} Q\left(s', a'; \theta_i^-\right)$$

Fig. 2: Target value [19]

As the size of the action space is not constant, a masking technique was used for the outputs of the network. It is a technique that masks the absent actions with the infinitely negative Q value, so the agent does not choose that action. It is completely differentiable in the calculation of the loss and the back-propagation. [20].

C. Dataset

The information in the dataset was gathered in a previous project [21]. The study was conducted during 10 sessions to track changes in the children's behaviour and engagement over time. The data comes from a rehabilitation program that involved 34 children with different levels of autism, participating in eight robot-assisted activities that varied in terms of social interaction. OpenPose was used for pose tracking and the engagement of the children was measured using both a binary system and a scale to assess their level of involvement.

IV. System Design

For this research, we utilized two types of robots: the humanoid NAO robot, which was manufactured by SoftBank Robotics, and the Furhat robot by Furhat Robotics.

A. NAO Robot

NAO is the primarily used robot in autism therapy, with nearly one-third of published studies in the past decade [18]. It is an autonomous and programmable robot often used for child-robot interaction research. It has basic modules such as built-in speech recognition, face recognition, display of gestures and body postures, and a text-to-speech engine that enables it to function more naturally.

In the NAO robot, we implemented 24 multi-purposeful activities which are divided into 8 categories. Each category is designed to train some type of subject or some type of skill. In every category there designed at least 3 activities. They are organized according to particular social skills such as joint attention, imitation, turn-taking, emotion labelling and display, and other core social skills.

Touch Me is a physical interactive game for children. During the game, children learn about body parts through the game. However, the main aim of this task is to teach a child to interact with someone like him, because most of the time autistic children are scared of physical interaction. The robot NAO may help to overcome this barrier.

Storytelling is an activity in which a robot tells fairy tales to the child and expresses emotions according to the story characters. We implemented several storytelling activities: ”Repka”, ”Kolobok”, ”Kyshik pen Mysyq”, and ”Zhyl Basy”. It is used to develop attention skills.

Social Acts is an interactive game where the child should repeat the common social acts, like greetings, and many others. The robot asks to press on one of the sensors, after pressing the sensor, the robot plays a random common social act and asks a child to name it and express it.

Emotions In this game robot encourages a child to recognize and express five different emotions. After the game starts robot randomly expresses one out of five emotions and asks ”What is this emotion?” and the child repeats it.

Songs is an activity where we implemented several songs with simple dances for them. List of available songs: ”Pauchok”, ”Chasiki”, ”Top Top Balaqan”, ”Azhe Shai”, ”Fixiki”, ”Aigolek”, ”Kap Kap Tuk Tuk”, and ”Mama”.

Dances is an active interaction game where the child is supposed to dance with the robot. List of available dances: ”Macarena”, and ”Gangnam Style”.

Imitations. In this activity, a child is supposed to imitate and recognize the actions that the robot performed. List of available imitations: ”Animals”, ”Transports”, and ”Sports”. In ”Animals” the child imitates different animals, in ”sports” the child repeats after the robot some sports actions, and in ”Transport” the child sound-imitates noises associated with one or the other transport.

Hello and Bye. This is a simple greeting from a robot to a child. The robot says: ”Hello, my name is NAO! Let's play” and at the end of the activity robot says ”It was nice to play with you! Good Bye!”

B. Graphical User Interface for the NAO robot

The goal of the NAO robot's Graphical User Interface (GUI) is to offer a straightforward and accessible interface for users to interact with the robot and manage its different capabilities. PyQt5 serves as the foundation for the GUI, with the primary User Interface (UI) files designed utilizing Qt Creator, and the system operating on Python 3.9. The GUI functions as a launchpad for a variety of applications sorted into seven primary categories and a database.

Database is needed for the therapist to know what kind of activities the child prefers and which one is not. Also, a database is used for collecting log data from each session, which is used as a source for analysis. The GUI is containing a therapist's comments section, the purpose of that section in a GUI for therapy software is to provide a space for the therapist to document their observations, assessments, and other important information related to the therapy session. This information can include the client's progress, challenges, goals, and any other relevant details that the therapist wants to note down. This information can be used to track the client's progress over time, make adjustments to the treatment plan as needed, and share information with other healthcare professionals involved in the client's care.

C. Furhat Robot Applications

Furhat robot is a social conversational robot that can communicate with people by imitating how people interact with each other. The robot consists of a facial mask, that allows one to show different faces and mimic emotions, a FOV camera, a stereo microphone and a stereo system. Robot has a natural language processing and internal library. Also, it could easily be integrated with other applications and software.

On the Furhat robot, we have implemented 20 applications for different aged groups. The applications aimed to interact with a child via interactive stories and games, where the robot and child work in tandem in order to go through the stories. The ability to change the person on the robot and change the voice allows the creation of complex multi-agent interactions. Applications are divided into several groups, such as Social stories, Interactive stories and Speech therapies. All applications are implemented in two languages Kazakh and Russian. To generate audio speech text to Speech model from the Human-Robot interaction Lab was integrated with the Furhat robot. The model is capable of generating 3 different Kazakh voices: man, woman and child voice. Also, we have integrated the Furhat robot with OpenAI API, in order to implement advanced adapting conversational agents. The robot receives the information in the Kazakh language using the automatic speech recognition system, then input is translated to English using machine translation since the language model better works with the English language and sends it to the language model. The output of the model then translates to Kazakh and voiced by TTS model, Figure 3.

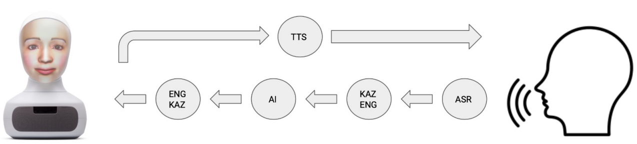


Fig. 3: Furhat pipeline

V. Results

A. Reinforcement-learning

The study investigated the performance of four different versions of a reinforcement learning model. To evaluate the performance of the models, both Returns and Mean Squared Error (MSE) Loss were measured. The first version used a state space that consisted of both non-binary engagement level (from 0 to 5) and Open-Pose data, Figure 4. The second version used only binary engagement level as the state space, Figure 5, while the third version used binary engagement level with Open-Pose data, Figure 6. The fourth and final version used only engagement level as the state space, Figure 7.

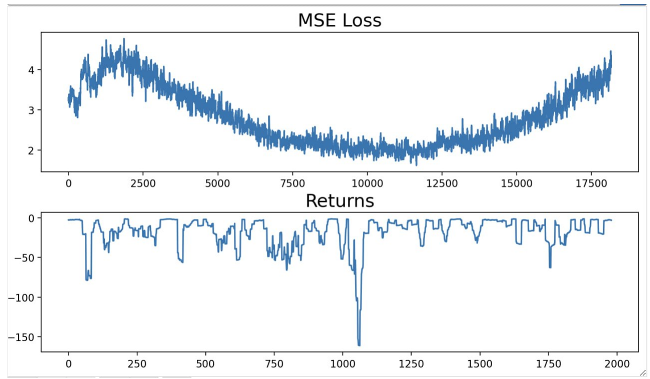


Fig. 4: Non-binary with poses

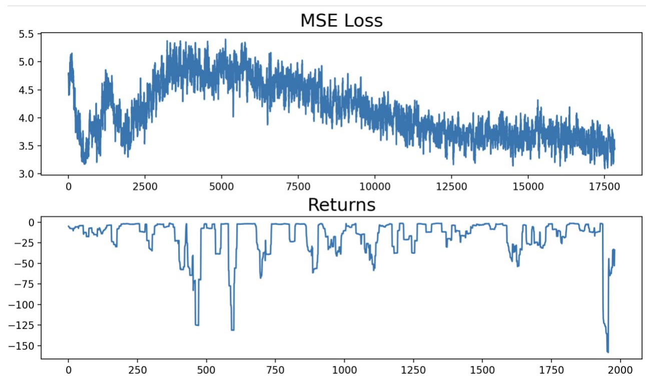


Fig. 5: Non-binary no poses

Overall, the results showed that all versions of the model have similar trends for the MSE Loss and returns with sudden drops in return pattern. That could be considered poorly trained models. However, the first and second versions of the model, which used non-binary engagement levels, had significantly higher MSE loss compared to the models with binary engagement. The third version, which used binary engagement level with Open-Pose data, had the lowest MSE loss and the highest returns. The fourth version, which used only binary engagement level as the state space, had similar MSE Loss as the third version, and lower returns.

Comparing the four versions, it appears that the binary engagement with Open-Pose data, Figure 6, was better at finding actions that adapt to the characteristics and behaviour of the child. This can be observed from the decreasing loss and high returns, even in the sudden drops. There are two possible reasons for this. The first reason is that the binary classification of engagement is more precise than the non-binary classification. It is possible that it is more complicated to analyse the engagement level from 0 to 5 rather than with 1 and 0. That is why using it as a part of the reward calculation gave better results. It can be observed that the returns from the binary classification are higher than those with the non-binary. The second reason might be that it is better to feed the actual poses to the network than its estimate made by the expert (engagement level), as it has more various and accurate data points. Comparing it to the environment with the binary classification and with no poses from Figure 7, we can observe that although the loss functions have similar patterns, the binary one has lower loss and higher returns.

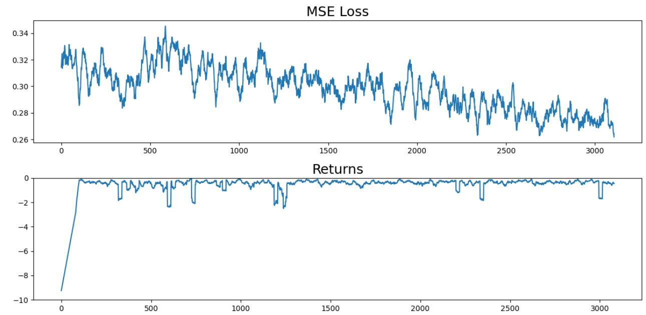


Fig. 6: Binary with poses

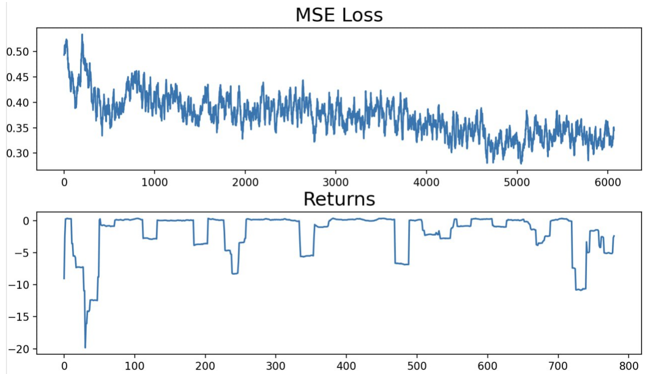


Fig. 7: Binary no poses

Also, the behaviour of the ASD children is unpredictable and does not follow expected patterns, this might have caused the inconsistencies in the engagement results in the dataset, which caused performance issues and sudden drops.

B. Autism center in Pavlodar

The implemented applications for the Furhat robot and NAO were used to present a RAAT at the Center for children with autism "Kamkorlyk" in Pavlodar. During the demonstration, we successfully showed and tested all NAO applications from 7 categories on children in autism centers, Figure 7.

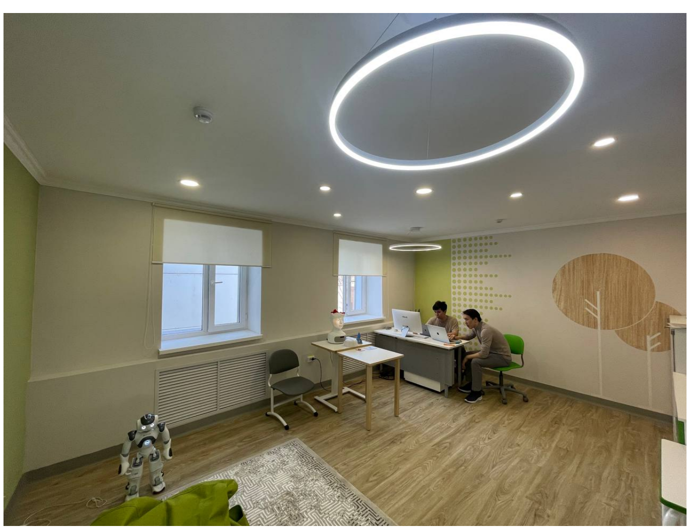


Fig. 8: Presentation in Pavlodar

Also, the proposed robotic therapy and applications are integrated within the project "Equipment of Centers for Children with Autism and other mental disorders", implemented by the "Kazakhstan Khalkyna Foundation" together with the "Kamkorlyk Kory" and "Samruk-Kazyna Trust" foundations. Within the framework of the project, 22 centers for children with autism spectrum disorders "Kamkorlyk" will be opened throughout the country. Our developed applications are part of the RAAT program in the centers.

VI. Conclusion

This study has demonstrated the potential of integrating open-source software and reinforcement learning algorithms in the development of adaptive robotic systems for robot-assisted autism therapy. The results have shown that a reinforcement learning algorithm could be implemented for RAAT. Also, results showed that models utilizing binary engagement levels for the state space were more effective in finding actions that adapted to the child's characteristics and behaviour. Our findings suggest that binary classification of engagement may yield better results than a non-binary classification for the reinforcement learning algorithm. Furthermore, utilizing actual poses as part of the input may provide a more accurate representation of the child's engagement level.

Implemented applications for the Furhat robot and NAO robots were successfully demonstrated and tested at the center for children with autism "Kamkorlyk" in Pavlodar. The applications are now part of the RAAT program in 22 centers across the country, highlighting the real-world impact of this research.

In conclusion, the integration of open-source software and reinforcement learning algorithms shows great promise in developing more sophisticated, adaptive robotic systems for robot-assisted therapy and advancing the field of Human-Robot Interaction.

References

[1] C. for Disease Control and Prevention, "Autism spectrum disorder," 2022. [Online]. Available: https://www.cdc.gov/ncbddd/autism/index.html

[2] M. Somerton, V. Stolyarova, and S. Khanin, "Autism and the knowledge and beliefs of specialists in kazakhstan," *Journal of Autism and Developmental Disorders*, Apr. 2021.

[3] J. Leaf, J. Cihon, J. Ferguson, and S. Weinkauf, *An Introduction to Applied Behavior Analysis*, 2018, pp. 25–42.

[4] O. Rudovic, J. Lee, M. Dai, B. Schuller, and R. W. Picard, "Personalized machine learning for robot perception of affect and engagement in autism therapy," *Science Robotics*, vol. 3, 2018.

[5] M. Chen, W. Xiao, L. Hu, Y. Ma, Y. Zhang, and G. Tao, "Cognitive wearable robotics for autism perception enhancement," *ACM Trans. Internet Technol.*, vol. 21, no. 4, 2021. [Online]. Available: https://doi.org/10.1145/3450630

[6] B. Scassellati, L. Boccanfuso, C.-M. Huang, M. Mademtzi, M. Qin, N. Salomons, P. Ventola, and F. Shic, "Improving social skills in children with asd using a long-term, in-home social robot," *Science Robotics*, vol. 3, 2018.

[7] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[8] M. Stolarz, A. Mitrevski, M. Wasil, and P.-G. Plöger, "Personalized behaviour models: A survey focusing on autism therapy applications," *ArXiv*, vol. abs/2205.08975, 2022.

[9] F. Zini, F. Le Piane, and M. Gaspari, "Adaptive cognitive training with reinforcement learning," *ACM Trans. Interact. Intell. Syst.*, vol. 12, no. 1, mar 2022. [Online]. Available: https://doi.org/10.1145/3476777

[10] D. Shawky and A. Badawi, "Towards a personalized learning experience using reinforcement learning," *Machine learning paradigms: Theory and application*, pp. 169–187, 2019.

[11] N. Akalin and A. Loutfi, “Reinforcement learning approaches in social robotics,” *Sensors*, vol. 21, no. 4, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/4/1292

[12] H. W. Park, I. Grover, S. Spaulding, L. Gomez, and C. Breazeal, “A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 687–694.

[13] S. Ali, F. Mehmood, Y. Ayaz, M. J. Khan, H. Sadia, and R. Nawaz, “Comparing the effectiveness of different reinforcement stimuli in a robotic therapy for children with asd,” *IEEE Access*, vol. 8, pp. 13 128–13 137, 2020.

[14] I. Salhi, M. Qbadou, S. Gouraguine, K. Mansouri, C. Lytridis, and V. Kaburlasos, “Towards robot-assisted therapy for children with autism—the ontological knowledge models and reinforcement learning-based algorithms,” *Frontiers in Robotics and AI*, vol. 9, 2022.

[15] R. Pakkar, C. Clabaugh, R. Lee, E. Deng, and M. Mataric, “Designing a socially assistive robot for long-term in-home use for children with autism spectrum disorders,” 01 2020.

[16] O. Rudovic, M. Zhang, B. Schuller, and R. Picard, “Multi-modal active learning from human data: A deep reinforcement learning approach,” *International Conference on Multimodal Interaction*, 14-18 October 2019.

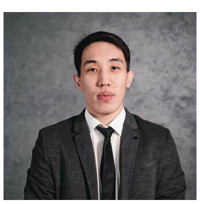
[17] C. E. Clabaugh, K. Mahajan, S. Jain, R. Pakkar, D. Becerra, Z. Shi, E. Deng, R. Lee, G. Ragusa, and M. J. Matarić, “Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders,” *Frontiers in Robotics and AI*, vol. 6, 2019.

[18] M. Saleh, F. Hanapiah, and H. Hashim, “Robot applications for autism: a comprehensive review,” *Disability and Rehabilitation: Assistive Technology*, vol. 16, pp. 1–23, 07 2020.

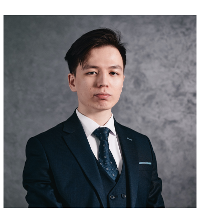
[19] F. Tan, P. Yan, and X. Guan, “Deep reinforcement learning: From q-learning to deep q-learning,” in *Neural Information Processing*, D. Liu, S. Xie, Y. Li, D. Zhao, and E.-S. M. El-Alfy, Eds. Cham: Springer International Publishing, 2017, pp. 475–483.

[20] I. Grondman, L. Busoniu, G. A. Lopes, and R. Babuska, “A survey of actor-critic reinforcement learning: Standard and natural policy gradients,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1291–1307, 2012.

[21] A. Zhanatkyzy, Z. Telisheva, A. Amirova, N. Rakhymbayeva, and A. Sandygulova, “Multi-purposeful activities for robot-assisted autism therapy: What works best for children’s social outcomes?” 03 2023, pp. 34–43.



Ilyas Issa is a senior Robotics and Mechatronics student. Integrated Open AI ChatGPT, Kazakh Speech Recognition and Kazakh TTS into the Furhat robot. Prepared interaction models for Furhat Robot. Assist in debugging and writing code for an environment for the RL and DQN algorithm.



Saparkhan Kassymbekov is a senior Robotics and Mechatronics student. Implemented an environment for the RL agent and coded the DQN algorithm. Prepared the available dataset and the model for NAO to work on the RL environment.

Zholaman Kuangaliev is a senior Robotics and Mechatronics student. Implemented all applications for the NAO robot. Working on adding engagement recognition for the NAO robot to work with the RL and on integrating Kazakh TTS into the NAO robot.

Symbat Nurgazy is a senior Robotics and Mechatronics student. She created and translated materials into Kazakh and Russian for the NAO robot tasks. Prepared interaction models for Furhat Robot. Researched autism therapy and the needs of children on the spectrum.