ENGAGEMENT RECOGNITION WITHIN ROBOT-ASSISTED AUTISM THERAPY

NAZERKE
RAKHYMBAYEVA

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Science, Engineering and Technology

School of Engineering and Digital Sciences
Nazarbayev University

Supervised by
Lead supervisor: Associate Professor, Anara Sandygulova
Internal co-supervisor: Associate Professor, Amin Zollanvari
External co-supervisors:
Professor at Ghent University, Tony Belpaeme
Senior Lecturer at the University of Melbourne, Wafa Johal

2023

Declaration

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the author's original work. The thesis has not been previously submitted to this or any other university for a degree and does not incorporate any material already submitted for a degree.

Signed 

Dated 26/10/2023

ENGAGEMENT RECOGNITION WITHIN ROBOT-ASSISTED AUTISM THERAPY

Nazerke Rakhymbayeva

Abstract

Autism is a neurodevelopmental condition typically diagnosed in early childhood, which is characterized by challenges in using language and understanding abstract concepts, effective communication, and building social relationships.

The utilization of social robots in autism therapy represents a significant area of research. An increasing number of studies explore the use of social robots as mediators between therapists and children diagnosed with autism. Assessing a child's engagement can enhance the effectiveness of robot-assisted interventions while also providing an objective metric for later analysis.

The thesis begins with a comprehensive multiple-session study involving 11 children diagnosed with autism and Attention Deficit Hyperactivity Disorder (ADHD). This study employs multi-purposeful robot activities designed to target various aspects of autism. The study yields both quantitative and qualitative findings based on four behavioural measures that were obtained from video recordings of the sessions. Statistical analysis reveals that adaptive therapy provides a longer engagement duration as compared to non-adaptive therapy sessions. Engagement is a key element in evaluating autism therapy sessions that are needed for acquiring knowledge and practising new skills necessary for social and cognitive development.

With the aim to create an engagement recognition model, this research work also involves the manual labelling of collected videos to generate a QAMQOR dataset. This dataset comprises 194 therapy sessions, spanning over 48 hours of video recordings. Additionally, it includes demographic information for 34 children diagnosed with ASD. It is important to note that videos of 23 children with autism were collected from previous records. The QAMQOR dataset was evaluated using standard machine learning and deep learning approaches. However, the development of an ac-

curate engagement recognition model remains challenging due to the unique personal characteristics of each individual with autism. In order to address this challenge and improve recognition accuracy, this PhD work also explores a data-driven model using transfer learning techniques. Our study contributes to addressing the challenges faced by machine learning in recognizing engagement among children with autism, such as diverse engagement activities, multimodal raw data, and the resources and time required for data collection.

This research work contributes to the growing field of using social robots in autism therapy by illuminating an understanding of the importance of adaptive therapy and providing valuable insights into engagement recognition. The findings serve as a foundation for further advancements in personalized and effective robot-assisted interventions for individuals with autism.

Publications

The part of the research conducted within this thesis has resulted in the following publications:

1. N.Rakhymbayeva, N.Seitkazina, D.Turabayev, A.Pak, and A.Sandygulova. 2020. A Long-term Study of Robot-Assisted Therapy for Children with Severe Autism and ADHD. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI'20). Association for Computing Machinery, New York, NY, USA, 401–402. https://doi.org/10.1145/3371382.3378356

2. N.Rakhymbayeva, A.Amirova and A.Sandygulova (2021) A Long-Term Engagement with a Social Robot for Autism Therapy. Frontiers in Robotics and AI 8:669972. doi: 10.3389/frobt.2021.669972

3. N.Rakhymbayeva and A.Sandygulova. Transfer Learning of Engagement Recognition within Robot-Assisted Therapy for Children with Autism. In: Proceedings of the AAAI Conference on Artificial Intelligence 35.18 (2021), pp. 15728–15729.

4. N.Rakhymbayeva, Z.Balgabekova, M.Nurmukhamed, K.Burunchina, W.Johal, and A.Sandygulova. 2022. To Transfer or Not To Transfer: Engagement Recognition within Robot-Assisted Autism Therapy. In Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI'22). IEEE Press, 1002–1006.

Some ideas and figures have appeared previously in the following publications as a co-author:

1. Z.Telisheva, A.Zhanatkyzy, A.Turarova, **N.Rakhymbayeva**, and A.Sandygulova. 2020. Automatic Engagement Recognition of Children within Robot-Mediated Autism Therapy. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI'20). Association for Computing Machinery, New York, NY, USA, 471–472. https://doi.org/10.1145/3371382.3378390

2. A.Amirova, **N.Rakhymbayeva**, E.Yadollahi, A.Sandygulova and W.Johal (2021) 10 Years of Human-NAO Interaction Research: A Scoping Review. Frontiers in Robotics and AI 8:744526. doi: 10.3389/frobt.2021.744526

3. A.Amirova, **N.Rakhymbayeva**, A.Zhanatkyzy, Z.Telisheva, and A.Sandygulova. (2022). Effects of Parental Involvement in Robot-Assisted Autism Therapy. Journal of Autism and Developmental Disorders 53, 438–455 (2023). https://doi.org/10.1007/s10803-022-05429-x

4. Z.Telisheva, A.Amirova, **N.Rakhymbayeva**, A.Zhanatkyzy, A.Sandygulova. The Quantitative Case-by-Case Analyses of the Socio-Emotional Outcomes of Children with ASD in Robot-Assisted Autism Therapy. Multimodal Technologies and Interaction. 2022, 6, 46. https://doi.org/10.3390/mti6060046

5. A.Sandygulova, A.Amirova, Z.Telisheva, A.Zhanatkyzy, and **N.Rakhymbayeva**. 2022. Individual Differences of Children with Autism in Robot-assisted Autism Therapy. In Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI'22). IEEE Press, 43–52.

6. A.Zhanatkyzy, Z.Telisheva, A.Amirova, **N.Rakhymbayeva**, and A.Sandygulova. 2023. Multi-Purposeful Activities for Robot-Assisted Autism Therapy: What Works Best for Children's Social Outcomes? In Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI'23). Association for Computing Machinery, New York, NY, USA, 34–43. https://doi.org/10.1145/3568162.3576963

Acknowledgments

I would like to express my sincere appreciation to my thesis supervisor, Anara Sandygulova, for her invaluable support and guidance during this research project. Her knowledge and expertise in the HRI field have been instrumental in shaping the direction of this thesis, and her feedback and persistence have been pivotal in achieving its success. I am grateful for the support and encouragement of my supervisor, who believed in my potential from the outset.

Additionally, I want to say thank you to the Department of Robotics and Mechatronics in the School of Engineering and Digital Sciences at Nazarbayev University (NU) for providing me with the opportunity to pursue my research interests. Working in the HRI lab has been a fulfilling and enriching experience.

I am also indebted to the numerous individuals who have contributed to this research project. The members of the HRI-lab at NU, including Aida Zhanatkyzy, Zhansaule Telisheva, and Aizada Turarova invaluable assistance in collecting and labelling the QAMQOR dataset's engagement level. And another member Aida Amirova for her support in reporting the objectives for every nine blocks of activities implemented for the interaction with the robot. I would also like to acknowledge the contributions of bachelor students of NU, including Nurila Seitkazina, Dauren Turabayev, and Alina Pak, who provided crucial assistance in collecting data from the third cohort of children and designing and implementing the "Social Acts" activity. The contributions of bachelor students Mukhamedzhan Nurmukhamed, Zarema Balgabekova, and Karina Burunchina in transfer learning from the PInSoRo dataset to the QAMQOR dataset have also been instrumental. I would also like to thank graduate students, Kenesary Koishybayev for his technical expertise and support in implementing a simple model for RNN and Aidyn Ubinguzhinov for helping me to extract MediaPipe keypoints from video frames for the QAMQOR dataset. Their hard work and dedication have been critical in the success of this research project, and I am deeply grateful for their contributions.

Furthermore, I would like to thank the participants and their parents who took

part in the robot-assisted therapy sessions, as well as the staff of the Rehabilitation Center for their assistance in conducting this research.

During my time at NU, I have been fortunate to have many mentors, and I am deeply grateful for their guidance and support. I would like to thank Professors Amin Zollanvari, Behrouz Maham, Luis Rojas, and Konstantinos Kostas for their valuable feedback and insights. I would also like to extend my thanks to my external co-supervisors, Prof.Tony Belpaeme and Prof.Wafa Johal, for their invaluable input.

Last, but not least, I want to express my gratitude to my family for their unwavering support and encouragement throughout my academic journey. Their belief in my potential has been a constant source of inspiration, and I am grateful for their sacrifices and dedication.

Contents

List of Tables

List of Figures

List of Acronyms

ABA Applied Behavior Analysis

ASD Autism Spectrum Disorder

ADHD Attention Deficit Hyperactivity Disorder

CRI Child-Robot Interaction

HRI Human-Robot Interaction

RAT Robot-Assisted Therapy

RAAT Robot-Assisted Autism Therapy

SAR socially Assistive Robotic

Chapter 1

Introduction

This chapter introduces the motivation behind this thesis research, objectives, research questions, hypotheses, challenges, and expected contributions.

1.1 What We Know About Autism

Autism Spectrum Disorder (ASD) is a condition that affects neurodevelopment and is marked by persistent and significant deficits in interests, activities, or social communication, as well as repetitive behaviours [1, 11]. Early signs of autism include a lack of interest in other people and limited eye contact. This description is based on the criteria set by the Centers for Disease Control and Prevention [1] and the American Psychiatric Association in 2013 [11]. However, the effects of autism can vary widely between individuals. Like neurotypical children, children with autism have unique strengths and weaknesses. For instance, they may struggle with basic social communication and interaction, exhibit repetitive and restrictive behaviours, and be oversensitive or undersensitive to sensory stimuli such as light, colours, sounds, temperatures, pain, touch, tastes, or smells. Additionally, regulating emotions can be challenging [14]. Despite these challenges, individuals with autism can demonstrate exceptional focus and expertise in their hobbies and interests, which can translate into academic and professional success in their fields of interest.

Autism is a condition that affects all human groups, regardless of their nationality,

social class, or culture, and has been identified across all racial, ethnic, and social demographic groups [132]. However, according to Maenner et al. [132], the incidence of autism diagnosis is four times higher in boys than in girls. Also, ASD affects approximately 26 in every 100 000 children globally [234]. However, tracking the global prevalence of ASD is challenging, as it varies significantly from one country to another, as shown in Figure 1-1 [234]. In Kazakhstan, for instance, the number of diagnosed cases of autism in children has increased by a factor of eight over the past eight years, as shown in Figure 1-2 [12].

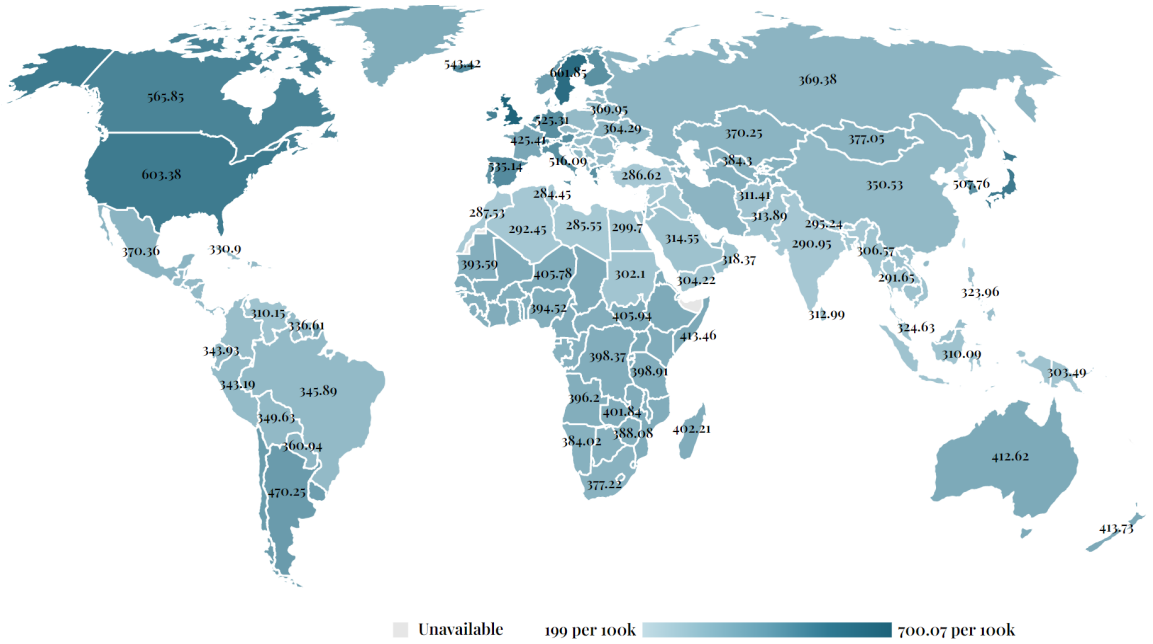

Figure 1-1: Autism rates by country [234]

Diagnosing a child with ASD requires a multidisciplinary evaluation that involves several doctors and specialists, including a paediatrician, psychologist, speech pathologist, psychiatrist, and occupational therapist [4]. However, assessing and rehabilitating children with autism can be challenging for even the most experienced specialists, given the wide range of symptoms and the fact that autism affects each child differently [11, 14, 113, 134]. As Dr Stephen Shore, an autism advocate said, "If you've met one person with autism, you've met one person with autism" [133]. Each child with ASD sees and senses the environment in a unique way [131], and their responses to social situations may differ depending on the severity of their condition, which is

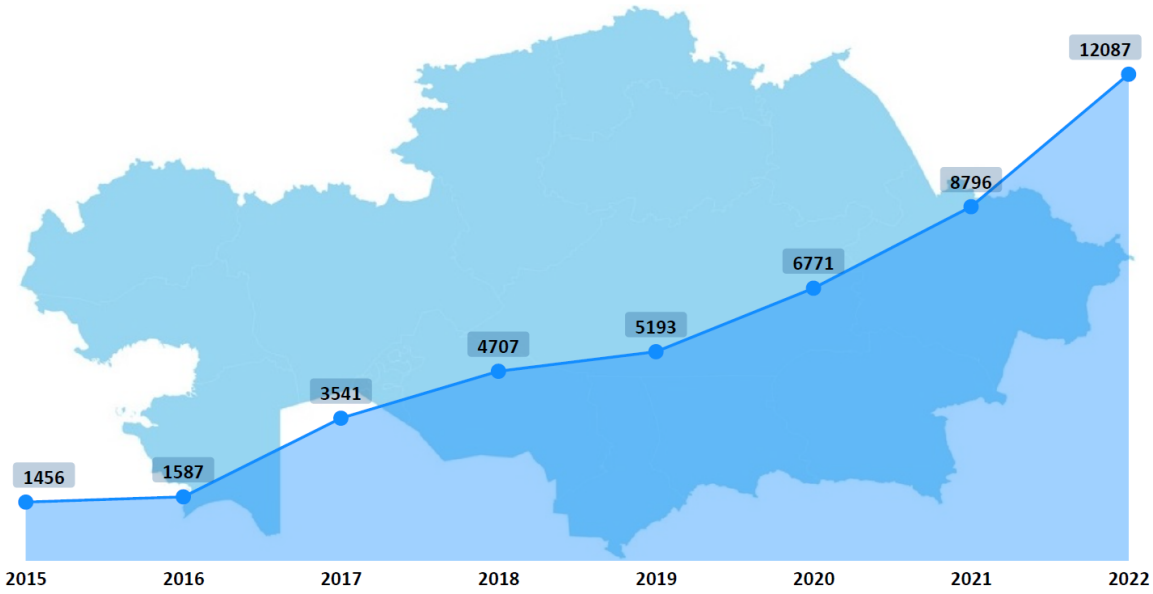

Figure 1-2: ASD prevalence in Kazakhstan: the number of identified cases

often classified as mild, moderate, or severe [11]. Therefore, therapy should include different approaches tailored to each individual, such as behavioural, educational, psychological, social-relational, developmental, and pharmacological therapies [93] (see Table 1.1 for more information). These therapies aim to strengthen the improvement of social, emotional and life skills. Given the variability of the autism spectrum, it is critical to continue exploring ways to personalize therapy for each child and scale it to provide the most effective outcomes.

Researchers across various fields are exploring different ways to work with children diagnosed with ASD, including the use of technology. Golestan et al. [87] conducted a comprehensive review of the types of technologies used for the diagnosis, assessment, and rehabilitation of children with autism. They categorized technologies into two groups: software-based, which is focused on software development, and hardware-based, which involves designing and using devices. The authors further divided hardware-based technologies into two subgroups: robots and dedicated devices [87]. Robots have been primarily used to address social and communication skill difficulties, often mediating between children and therapists.

Table 1.1: Types of therapy approaches for children diagnosed with ASD [93]

#	**Approaches**	**Focus**	**Examples**
1	Behavioral	changing behaviours by understanding what happens before and after the behaviour	Applied Behavioral Analysis (ABA), Discrete Trial Training (DTT), Pivotal Response Training (PRT)
2	Educational	given in the classroom settings	Treatment and Education of Autistic and Related Communication-Handicapped Children (TEACCH), Emotional Regulation (SCERTS), Social Communication
3	Psychological	helping to cope with depression, anxiety, and similar mental health issues	Cognitive-Behavior Therapy (CBT)
4	Social-Relational	improving social skills and emotional bonds	Relationship Development Intervention (RDI), Relationship-Based, Developmental, Social Skills Groups, Individual Differences, Social Stories
5	Developmental	improving specific developmental skills	Speech and Early Start Denver Model, Sensory Integration Therapy, Physical Therapy, Language Therapy
6	Pharmacological	helping to manage high energy levels, inability to focus, or self-harming behaviour	Medications

1.2 Employing Social Robots in Autism Therapy

The use of robots in healthcare settings, particularly in therapy and pedagogical interventions with children, is increasing. Social robots, with their tangible 3D presence and being able to show complex behaviour while looking safer than human beings, bridge a gap from the real to the virtual world and otherwise. One area which is gaining significant interest from both researchers and the public is employing robots in autism therapy [214, 215].

Research has shown that the use of robots is well-accepted and can lead to progressive improvements in communication complexity, based on an individual's needs and abilities. This innovative application of robots is especially promising as it has the ability to get beyond some of the drawbacks of traditional therapeutic techniques. With further research and development, robots could become an important tool in

improving the quality of support for individuals with autism and other developmental disorders.

Robots have shown promising results in robot-assisted therapy (RAT) for children diagnosed with ASD [96, 151, 181, 205, 220, 238]. Studies have shown that robots can positively impact various areas of development, including imitation skills [26, 68, 88, 199, 210], emotion recognition [46, 125, 157, 201], verbal and nonverbal communication [108, 112], eye contact [9, 34, 114], joint attention [9, 8, 121, 186, 189, 207], simple activity sharing [153], self-initiated interaction [58], and behavioural response [66] given various robot capabilities.

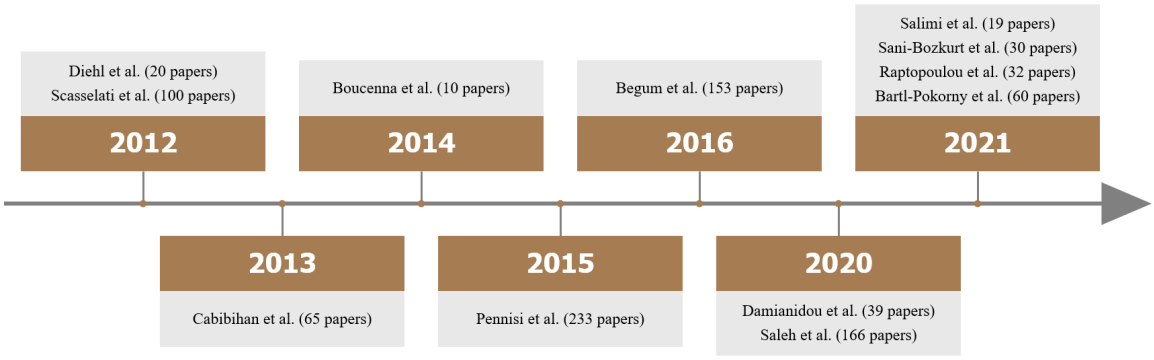

Figure 1-3: The publication year of the survey and systematic review papers

Several literature reviews discussed the use of social robots to assist people with ASD (Figure 1-3). Diehl et al. [64] presented a comprehensive review of the clinical use of robots up to March 2011, where they classified the utility of robots into four categories:

- child's perception of a robot (with no direct clinical application);
- the robot's ability to extract the target behaviour in a child;
- the robot (as a teacher, agent or model) teaching the child to perform a specific skill;
- the robot accommodating feedback and encouragement to the children.

These categories provide a framework for understanding the various ways that robots

can assist in autism therapy, and highlight the importance of considering individual needs and abilities when implementing RAT.

In the review by Scassellati et al. in 2012 [187], it is concluded that social robots can serve as effective tools for therapy, serving two primary roles: a leader and a mediator, and even acting as a proxy for children. Scassellati et al. [187] suggest that one of the advantages of using robots in therapy is their consistency and predictability, which can help foster a sense of tranquillity and trust in children. Additionally, because robots do not experience emotional fatigue or disappointment, they may be able to provide a stable and reliable source of support for children with autism in a way that human therapists may not be able to. However, despite the potential benefits of using social robots in therapy, the review also found that there is a dearth of long-term studies exploring their efficacy, a need for more rigorous experimental protocols and a dependence on the Wizard of Oz (WoZ) setup to control the behaviour of the robots.

In addition, Cabibihan et al. [31] conducted a review in 2013 that provided a technical perspective on social robots used in ASD intervention. This review highlighted three key aspects:

- Robot design attributes, which included autonomy, functionality, and embodiment, were important considerations for the effective use of robots in therapy.

- The role of the robot in therapy sessions could vary depending on the specific goals of the intervention. Robots could serve as diagnostic agents, social mediators, or peers, among other roles.

- The robot's behaviour during therapy sessions was a critical factor in determining its effectiveness. The review emphasized the importance of robot behaviours that foster engagement, communication, and social interaction with children diagnosed with ASD.

The review paper by Boucenna et al. in 2014 [27] explored the potential of interactive communication technologies for children with autism. The authors found

that interactive devices could be utilized in a number of ways, including evaluating an action of the children towards the robot behaviours, detecting and responding to a child's behaviours, teaching and practising skills, and providing feedback on the interaction. Despite the promise of these devices, the authors highlight the need for further research on the development of autonomous robots with human-like abilities, such as joint attention, interactive engagement, and imitation, to provide effective therapies for early development in children with autism [27].

Pennisi et al. (2015) [151] reviewed the potential of social robots as mediators and attractors for individuals with ASD with the aim of reducing repetitive and stereotyped behaviours. The authors suggest that while social robots can be effective in this regard, they may also act as distractors during some therapy sessions [151].

In 2016's review paper, Begum et al. [17] investigated the current state of RAT studies. The authors found that many studies had small sample sizes, did not use standard research designs, and relied on custom-made therapies [17].

A more recent review by Damianidou et al. in 2020 [52] examined the employing of social robots as behaviour-eliciting agents to develop social, communication and interaction skills, such as verbal initiations, joint attention, eye contact, and gesture production and recognition. The authors highlighted that social robots showed promising results in helping children with ASD develop these skills [52].

Saleh et al. (2020) [181] presented a comprehensive review of studies that employed social robots for the diagnosis, rehabilitation, and education of children with autism. The authors categorized the studies into ten groups based on their goals and found that the NAO robot was frequently used in studies aimed at improving learning skills [181].

In order to find out whether the robot is better than traditional methods, Salimi et al. [182] conducted a systematic review in 2021. The authors highlighted that the engagement of children rises with robots, but it might decrease when children get used to it. Moreover, social robots were used to teach gesture production and recognition and to help prepare individuals for a job [182].

A systematic review of studies that examined the effects of social robots on im-

proving joint attention in children with ASD was conducted by Sani-Bozkurt et al. (2021) [186]. Their review analyzed a total of 11 studies, which collectively involved 166 children with ASD. The results of the review showed that social robots have a positive effect on improving joint attention in children with ASD. Specifically, the studies found that the use of social robots led to increased joint attention behaviours such as gaze following, pointing, and shared attention. The review also found that the use of social robots during therapy sessions resulted in increased engagement and motivation in children with ASD, as well as decreased levels of negative behaviours.

Raptopoulou et al. (2021) [160] reported a review of works that used social robots to improve communication and social skills. Based on their analysis, most of the studies used anthropomorphic robots. Also, many studies did not have a control group and did not present follow-up sessions [160]. Similar to the findings of Begum et al. [17], these studies also had a limited number of samples.

Bartl-Pokorny et al. [15] conducted a systematic literature review of RAT studies in 2021, with a focus on emotion recognition and interaction skills. The authors noted that many reviewed papers involved small sample sizes, and identified a gap in research regarding factors that may affect child-robot interaction (CRI) [15].

Overall, several literature reviews have discussed the use of social robots for therapy in individuals diagnosed with ASD. These reviews suggest that social robots can be effective tools for therapy by providing consistency and predictability, accommodating feedback and encouragement to children, and serving as leaders and mediators [187]. However, more rigorous experimental protocols and long-term studies exploring efficacy are needed, as well as the development of autonomous robots with human-like abilities for effective early development therapies. Social robots were found to improve interaction skills, reduce repetitive and stereotyped behaviours, enhance learning skills, and improve joint attention in children with ASD. However, studies with a lack of standard research designs, custom-made therapies, small sample sizes, and the potential for social robots to act as distractors during therapy sessions have been identified as limitations. Overall, robots can be effective motivators and keep children engaged, thereby improving targeted skills. However, it is important to emphasize

that social robots are not intended to replace therapists.

More than 40 different robots have been utilized in RAT with the aim of improving social and communication skills, eye gaze, imitation, and other targeted skills, as reported in the above-mentioned review papers [15, 116]. For more information on the types of robots and their capabilities used in autism therapy, please refer to Section 2.2.

The use of social robots in ASD therapy and education is an area of study that is still growing. However, more research is required to fully understand their efficacy and to fully explore their possibilities in bigger and more thorough studies. The complex and diverse needs of children diagnosed with ASD must be taken into consideration in order to create social robot therapies that aimed to be successful. Measuring the engagement of participants during therapy sessions can provide helpful information for adapting robot behaviour and evaluating the effectiveness of therapy [47].

1.3 Definition of Engagement in HRI

Engagement is a crucial concept in human-robot interaction (HRI) because it refers to the level of interest, attention, and emotional connection that users feel toward the robot. However, the definition of engagement can be complex and multidimensional, which may pose challenges for researchers new to this field.

From a psychological perspective, engagement refers to a continuous state of emotional, cognitive, and behavioural activation in individuals [82]. Beyond this general definition, engagement in HRI has been found to encompass various aspects of behaviour, emotion, and cognition [80, 228].

Behavioural engagement, for instance, relates to the extent of a user's interaction or involvement with the robot, including their willingness to participate in activities or pursue goals [79, 203]. Emotional engagement, on the other hand, involves a user's affective reactions to the robot, such as their feelings of happiness, sadness, anxiety, or boredom during the interaction [79, 203]. Lastly, cognitive engagement refers to a user's psychological commitment to the interaction, such as their use of learning

strategies and self-regulation strategies [228]. Therefore, understanding the different dimensions of engagement is essential in the development and evaluation of effective HRI systems that promote positive user experiences and sustained interaction.

The article by Oertel et al. [143] provides an overview of engagement research in HRI, covering various aspects such as task-based and social engagement. This research can be classified into two categories: engagement as a process and engagement as a state [143]. Engagement as a process refers to the multiple processes that occur during an interaction, including starting, maintaining, and ending perceived connections [198]. Engagement as a state can be characterized as either engaged or not engaged [95].

Engagement in HRI is measured using both explicit and implicit measures [114]. Explicit measures, such as questionnaires, provide insight into participants' interests and interactions but have limitations, including relying on self-reporting and being affected by social desirability bias. In contrast, implicit measures, such as performance measures and neurophysiological/neuroimaging metrics [83], allow for the study of perception, cognition, and behaviour during interactions and provide more accurate data on cognitive processes without requiring participants to be aware of the procedures of the experiment.

Oertel et al. [143] also differentiate studies based on whether they focus on the perception of an engagement or the generation of engagement-relevant behaviours. Researchers employ two approaches, rule-based [86, 162] and machine learning-based [37, 185], to measure engagement. Rule-based approaches involve explicitly defining engagement-relevant behaviours, whereas machine learning-based approaches use algorithms to identify relevant behaviours automatically. Machine learning-based approaches include reinforcement learning [178] and deep learning [7, 155].

Overall, the methodology for measuring engagement is crucial for quantification and generalizability. However, annotating engagement is a challenging task due to various definitions of engagement and the lack of a widely used scheme.

1.4 Research Objectives

Recent advancements in computer vision research have enabled more accurate recognition of engagement with the availability of large datasets. In therapy, the ability of a robot to monitor a child's level of engagement might be useful to provide adaptive interventions.

There are some challenges associated with collecting and annotating data in this field, including the time-consuming nature of video and/or audio recording of sessions [176] and the need for expert knowledge and experience [126].

The literature review has highlighted the need for an adaptive approach that caters to the individual needs and differences of children with diverse forms of ASD, and that focuses on long-term engagement in rehabilitation interventions [43, 176]. Therefore, the goal of this PhD thesis is to provide an adaptive learning experience in robot-assisted autism therapy (RAAT) for children with ASD and ADHD and evaluate its effectiveness in improving cognitive and emotional well-being.

Next, we propose to take a data-driven approach to classify the engagement of children with autism in interactions with social robots or other humans. This approach has the potential to improve the robot's autonomy. We will employ a transfer learning approach by pre-training a model on a large CRI dataset to increase the QAMQOR dataset's engagement recognition accuracy. Moreover, to improve a trained model, we aim to use the idea of transfer learning on one activity for classification in the other activity data within the QAMQOR dataset.

1.5 Research Challenges

Engagement recognition of children with autism poses several challenges to machine learning, which are currently being explored by researchers:

- Engagement activities can vary widely among children with ASD and ADHD, as well as within the same child across different sessions. This heterogeneity requires classifiers that can handle inter- and intra-subject variability.

- The data consists of time series of spatial interactions, which may require investigating the temporal and spectral components in addition to the spatial features.

- The image data can be noisy due to various factors such as low-quality video frames, occlusions, and errors in feature extraction.

- Engagement is a multifaceted construct that involves multiple modalities, such as facial expressions, body movements, eye contact, and speech intonation. The raw data can have thousands of features per sample, making it challenging to extract meaningful information and reduce the dimensionality.

- Data collection is a resource-intensive process that involves recruiting and consenting participants, setting up the experiment, and collecting and labelling the data. It can be particularly challenging to obtain large and diverse datasets for machine learning, especially when the samples are human subjects with ASD and ADHD who may require special accommodations and ethical considerations.

To address these challenges and to increase the reliability, effectiveness, and generalizability of engagement recognition models, we suggest investigating a number of machine learning approaches and tactics, including multi-task learning, transfer learning, and data augmentation.

1.6 Research Hypotheses

The following four hypotheses were formulated in order to test our assumptions about the study.

H1: Increased scores of valence and engagement during interaction with the robot across several sessions will be the result of the adaptive sessions using activities that each child is familiar with in the multi-session study.

H2: Recognition of engagement would be more accurate if the model trains on multimodal data compared to the single type of data.

H3: A model pre-trained on a similar available CRI dataset would improve the engagement recognition accuracy of the QAMQOR dataset.

H4: We could transfer knowledge from one activity to another activity with similar instances (face and body features) using the transfer learning approach for engagement recognition.

1.7 Expected Contributions

This thesis makes significant contributions to various areas.

Contribution 1: In our long-term RAAT study with children with ASD and ADHD, we explored the effectiveness of adaptive experiences tailored to their unique needs. We observed significant behavioural improvements through sustained intervention engagement. By this research adds to the literature on building an adaptive approach based on ASD children's needs and differences, focusing on the long-term commitment to interventions.

Contribution 2: We created a multimodal QAMQOR dataset comprising 194 sessions and over 48 hours of video, along with demographic information for each of the 34 children diagnosed with ASD. This dataset helps address the data scarcity issue identified in the reviews by David et al. and Diehl et al., providing valuable resources for problem-solving in this field.

Contribution 3: To enhance the accuracy of engagement recognition in the QAMQOR dataset, we developed a data-driven model using transfer learning techniques. Our study contributes to addressing the challenges faced by machine learning in recognizing engagement among children with autism, such as diverse engagement activities,

multimodal raw data, and the resources and time required for data collection.

1.8 Thesis Structure

This thesis is organized into eight chapters, each with a specific focus.

- Chapter 1 introduces the problem of measuring engagement in children with autism and outlines the importance and contribution of the research to the fields of RAAT, social robotics, and CRI. It also provides descriptions of autism, social robots, and engagement in HRI.
- Chapter 2 gives an overview of the engagement research in RAAT. It includes a tentative definition of engagement, existing models of engagement recognition of children, past and current RAAT projects, social robots used in RAT studies, and available HRI datasets.
- Chapter 3 describes the activities (robot behaviours) implemented in the research for children with autism.
- Chapter 4 presents the experimental study conducted for data collection, including the design, settings, participants, measures, and procedure.
- Chapter 5 details the generation of the QAMQOR dataset, including information on the participants, data pre-processing steps, feature extraction as a measure of engagement, and the coding scheme.
- Chapter 6 provides the results of the QAMQOR dataset evaluation under different conditions using standard machine learning and neural network algorithms for engagement recognition.
- Chapter 7 focuses on the implementation of the transfer learning approach for engagement recognition.
- Chapter 8 summarizes the research thesis by highlighting the main results, the limitations of the research, and suggestions for future work.

Chapter 2

Literature Review

This chapter explores existing research studies relevant to the thesis, which is consisted of six sections. The chapter started by providing an overview of existing projects related to RAAT. In the second section, we examine the various types of social robots utilized in autism therapy. We also provide a detailed description of engagement in RAAT in the third section. The fourth section focuses on works that concentrate on existing models of engagement recognition during interactions in RAAT. In the fifth section, we discuss existing long-term interaction studies with social robots in the context of autism therapy. Lastly, we provide a comprehensive summary of some existing HRI datasets, that researchers can use to train social robots, including general descriptions of how they were collected and what features they have in the sixth section.

2.1 RAAT Projects

Previous research has demonstrated that social robots can serve as a mediator in conventional methods of behavioural interventions [18, 179, 235]. Many projects worldwide have utilized social robots with the aim of improving a number of social behaviours, including imitation, joint attention and eye contact. We present a comprehensive overview of various existing projects that were carried out with children diagnosed with ASD, including their objectives, the solutions developed, the robots

used, and the outcomes achieved.

2.1.1 AuRoRa Project

The AuRoRA (Autonomous Robotic Platform as a Remedial Tool for Children with Autism) project investigated the use of robots for children with autism. The project, initiated by Prof. K. Dautenhahn in 1998 [54], aimed to engage children diagnosed with autism, facilitate tactile interaction, develop their communication skills (including imitation and turn-taking), teach cause-and-effect, and recognize emotional expressions and gestures.

The AuRoRa project explored the use of robots to engage children with autism, enhance their communication skills, and promote social interactions. The project's use of humanoid robots like Robota [56, 171] and KASPAR [53, 168], as well as various other robotic platforms (Labo-1 [231], Pekee [55], Aibo [78], and IROMEC [74]) and interactive software [60], has demonstrated the potential of technology-based interventions in supporting children with autism.

Before discussing the limitations, it is essential to acknowledge the strengths of the AuRoRa project. The research's multi-faceted approach, employing various robotic platforms and interactive software, allows for a comprehensive exploration of different interventions [57, 77, 223, 224, 225]. The development of the CRI model as a design and assessment framework provided a structured approach to integrating robots as social mediators in therapy [94, 167, 231]. Additionally, the creation of interactive games and activities [169, 172, 230], along with learning algorithms [77], demonstrated efforts to customize robot-assisted therapy to individual children's needs [63, 170, 171]. The project's guidelines for researchers and practitioners also contributed to shaping best practices in RAAT [59].

Furthermore, the long-term research involving KASPAR and its positive impact on social and communicative abilities in children with autism highlighted the potential benefits of utilizing humanoid robots as therapeutic tools [53, 168]. The success of the Theatrical Robot approach and the investigation of diverse evaluation methods showcased the project's commitment to innovative and evidence-based practices [171].

However, a critical analysis of the project reveals some limitations and challenges that need to be addressed for the advancement of RAAT:

- One of the main limitations of the AuRoRa project is the utilization of *small sample sizes* in some studies [171, 226]. Limited participant numbers can undermine the generalizability of the results and may not account for the diversity within the autism spectrum.
- While the project showcased promising short-term outcomes, *the long-term effectiveness* of robot-assisted therapy remains uncertain.
- Autism is a spectrum disorder with significant *individual variability.*
- The impact of RAAT interventions can be influenced by *cultural and social factors.*
- *The lack of standardized assessment* metrics makes it challenging to compare the effectiveness of RAAT across different studies.
- While the project demonstrated success in controlled research environments, *integrating RAAT into real-world therapeutic settings* can be complex.

Overall, this project has valuable contributions in RAAT and holds the potential to transform how technology can be utilized to support individuals with autism.

2.1.2 IROMEC Project

The IROMEC (Interactive Robotic Social Mediators as Companions) project, spanning from 2006 to 2009, built upon the success of its predecessor, AuRoRa. The IROMEC project involved collaboration between eight research partners from diverse fields such as information and communications technology, robotics, psychology, and pedagogy, showcasing its multidisciplinary nature. The primary objective of the project was to investigate how robotic toys, like IROMEC, could encourage children with ASD to participate in various games, both individually and cooperatively [163].

The IROMEC robot's incorporation of various sensors and interactive features allows for versatile and personalized gameplay, promoting cognitive, sensory, emotional, and social development in children with ASD [218]. The interdisciplinary nature of the IROMEC project was particularly commendable, as it demonstrated how collaboration across diverse fields can lead to meaningful outcomes in special education.

The games developed in this project have shown significant benefits in the areas of interaction, cognitive and sensory development, emotional and social growth, and even movement functions for children who have serious physical impairments [74, 218].

Despite the significant advancements made by the IROMEC project, several limitations and challenges warrant attention for the future of RAAT:

- The development and implementation of sophisticated robotic platform IROMEC, can be *expensive*, making them less accessible to many schools and families.
- There is a challenge in ensuring that the skills learned during robot-assisted therapy can be generalized to real-world social interactions. Moreover, *personalized and adaptive interventions* may be necessary to cater to the diverse needs of children with ASD.
- *Longitudinal studies* are essential to assess the lasting impact of robot-assisted therapy.
- *Acceptance* of robots as therapeutic tools may vary among individuals with ASD and their families.

Overall, the IROMEC project demonstrated a successful approach to employing robots as mediators in autism therapy and education for children in the field of special education. The project's success can be attributed to effective collaboration between various disciplines. Therefore, this project serves as an inspiring example of how interdisciplinary collaboration drives innovation and achieves meaningful outcomes for the development of abilities of children with ASD.

2.1.3 BabyRobot Project

The European Union (EU) Horizon 2020 Programme supported the BabyRobot project from 2016 to 2019, which aimed to advance the field of robotics and develop more effective tools for improving HRI to benefit both children with autism and other forms of diverse needs. The primary goal was to create robots that are able to track and analyze children's behaviours through audio-visual signals, thereby establishing collaboration between consumers and healthcare application markets.

The BabyRobot project's emphasis on modelling HRI through a three-step process of sharing attention, establishing common ground, and forming shared goals highlighted the importance of understanding and fostering effective communication between robots and children [164]. The use of multiple robots and sensor technologies, such as the Kinects, allowed for comprehensive data collection and analysis, enabling the development of tailored games and scenarios for children with autism [156]. The implementation of reinforcement learning and active exploration algorithms to enhance robot adaptability to changes in HRI demonstrated progress in creating more dynamic and responsive robotic companions [222].

Additionally, the project's efforts to improve the collaboration skills of the KASPAR robot through intention reading and shared intentionality modules showcased a step towards robots actively participating in collaborative activities with children [236]. The creation of 11 new games and scenarios tailored to educational and therapeutic needs further expanded the potential applications of robot-assisted interventions for children with ASD.

While the BabyRobot project made significant contributions to the field of RAAT, there are several limitations and challenges that need to be addressed in future:

- The project's evaluation of the system's effectiveness was based on a relatively small sample size of 28 children aged from six to ten years old. It remains uncertain whether the outcomes observed in this study *can be generalized* to a broader range of children with autism of different ages and developmental

levels.

- The project's focus on short-term interactions and game-based scenarios leaves questions about *the long-term impact* of robot-assisted interventions.

- The evaluation of the system was conducted with robots programmed to speak the Greek language. Cultural and language differences may influence children's responses to robot interactions, highlighting the need for *cross-cultural evaluations.*

Overall, the findings from this project can serve as a foundation for the development of future technologies that can positively impact the field of special education.

2.1.4 DE-ENIGMA Project

The DE-ENIGMA[1] project was designed to develop the social imagination skills of children with autism by employing a comprehensive, context-sensitive, multimodal, and naturalistic HRI approach. The project's primary objectives were to develop and implement effective solutions that would promote emotional, subjective, and social well-being in children with autism, utilizing advanced robotic technology.

By utilizing the Zeno robot and advanced robotic technology, the project aimed to promote emotional, subjective, and social well-being in children with autism [40]. The integration of personalized perception affect network (PPA-net) and multimodal active learning techniques showcased innovative approaches to achieve meaningful CRI [176]. While the project achieved its objectives and contributed to the field of robot-assisted training, a critical analysis reveals certain limitations and areas for future development.

The DE-ENIGMA project made significant advancements in HRI by employing a three-layered structure, PPA-net, that incorporated contextual data and individual characteristics of each child. A total of 128 children (62 British and 66 Serbian), aged between 5 and 12 years, participated. The integration of audio-visual and autonomic

[1]https://de-enigma.eu/

physiological recordings in the multimodal model allowed accurate estimation of the child's moods during interactions. By achieving a 65% intra-class correlation for engagement estimation of multimodal data, the project demonstrated progress in developing sophisticated models to understand and respond to children's emotional states during interactions with the robot [176].

The project's focus on fostering social and emotional abilities in children with autism, such as recognizing emotions and facial expressions, addressed essential aspects of social development for these children. The creation of a multimodal dataset accessible to academic researchers worldwide further contributed to the field by enabling future research and replication of the project's findings.

Despite its successes, the DE-ENIGMA project has some limitations and areas that warrant attention for future research:

- While the project involved 128 children in total, dividing them into human-assisted and robot-assisted sessions might have resulted in relatively *small sample sizes* for each condition.
- The project's design involved 4-5 sessions per child, which may be considered *short-term evaluations*.
- While the project made significant progress in developing personalized learning models, the challenge lies in determining how well these models can be *generalized* to a broader population of children with autism.

Despite certain weaknesses and areas for future development, the DE-ENIGMA project was primarily focused on constructing robot-assisted training rather than creating a diagnostic tool. The training was created to foster social and emotional abilities in children with ASD, such as recognizing emotions and facial expressions. Through the use of advanced robotics and personalized learning strategies, the DE-ENIGMA project achieved its objectives and demonstrated the potential for enhanced learning and support for children with autism.

2.1.5 DREAM Project

The DREAM[2] (Development of Robot-Enhanced therapy for children with AutisM spectrum disorders) project was conducted between 2014 and 2019 and funded by EU. This project was focused to enhance joint attention, imitation and turn-taking skills in children with autism.

The DREAM project made significant progress in developing a sensory system capable of recording and interpreting the behaviour of children with autism. By equipping the intelligent space environmental room with various sensors, such as Microsoft Kinect and high-resolution cameras, researchers gathered extensive data for signal analysis [32, 191]. The development of computational models to assess children's behaviour and infer their psychological disposition based on gaze, body language, and speech demonstrated a sophisticated and innovative approach.

Additionally, the implementation of a semi-autonomous decision-making system allowed the robot (NAO) to respond to the child's behaviour and provide appropriate feedback. The cognitive control architecture, consisting of multiple sub-systems, allowed therapists to oversee multiple children and plan individualized interventions, reducing the therapist's workload and increasing the efficiency of the intervention [33].

The development of a DREAM-lite version, consisting of only a tablet and a robot, showcased an attempt to simplify the technical requirements and make the system more accessible for real-life scenarios [135]. Testing this version with 67 children diagnosed with autism during multiple sessions provided valuable insights into the practical application of robot-enhanced therapy. The DREAM-lite version represents a step towards bridging the gap between research and practical implementation in real-world therapeutic settings.

Despite its advancements, the DREAM project has some limitations and considerations to address:

[2]https://dream2020.github.io/DREAM/

- The study focused on a specific set of skills (joint attention, imitation, and turn-taking), which may limit the *generalizability* of the findings to other areas of social and communicative development in children with autism.

- The project's sample may have *lacked diversity* in terms of cultural backgrounds and age groups, which could impact the applicability of the results to a broader population.

- While the project made significant progress in developing personalized learning models, the challenge lies in determining how well these models can be *generalized* to a broader population of children with autism.

In conclusion, the DREAM project utilized supervised or shared autonomy, with the therapist still controlling the robot's actions. Consequently, three primary subsystems were developed within the robot software: one for detecting and interpreting perceptual input, one for analyzing the behaviour of the child, and a third for managing the behaviour of the robot.

2.1.6 MICHELANGELO Project

The MICHELANGELO project was funded under the EU Programme from 2011 to 2014. The aim of the project was to produce an affordable, patient-centred, home-based intervention for autism therapy that requires less therapeutic involvement and evaluation outside of the clinical setting [24].

One of the notable strengths of the MICHELANGELO project was the proposal of affordable technological solutions for autism therapy in a home setting, e.g., GOLIAH [25]. By integrating different sensors [84] (RGB-D sensors, cameras, microphones, wearable systems for electroencephalography (EEG) [39] and electrocardiogram (ECG)) integrated with humanoid robots [8, 27] to record the signals of the children's behaviour in the controlled environment [10]. This approach enabled continuous monitoring and data collection during interventions, facilitating the assessment of therapy effectiveness.

Additionally, the use of ICT-based solutions like virtual reality, serious games, and robotics demonstrated a forward-looking approach to designing interactive and engaging interventions for children with autism [8, 24, 27, 75]. The incorporation of Diffusion Tensor Imaging and fMRI algorithms by therapists to monitor brain connectivity during therapy was a notable advancement in using advanced imaging techniques to assess therapy outcomes [24, 75].

The MICHELANGELO project's focus on joint attention and imitation through games aligned with the therapeutic needs of children with autism [8]. These core social communication skills are often challenging for children with autism, and interventions targeting these areas have the potential to make a significant impact on their development.

Despite its advancements, the MICHELANGELO project has some limitations and considerations to address:

- The project's sample size, particularly the number of children with autism and typically developed children, may be considered relatively small [8].
- Implementing therapy in a home setting can introduce additional challenges, such as ensuring consistent and appropriate usage of the technological tools by caregivers and children [25].

The MICHELANGELO project made advancements in developing technological solutions for home-based autism therapy. Its focus on joint attention and imitation, along with the use of ICT-based interventions and advanced imaging techniques, demonstrated a forward-looking approach in autism research.

2.1.7 PicASSo Project

The PicASSo project started in 2014 in a collaboration between the Technical University Eindhoven, Karakter Centre for Child and Adolescent Psychiatry, and the Radboud University Medical Center. This project aimed to examine the efficacy of

employing the social robot NAO during parent-mediated Pivotal Response Treatment (PRT) to decrease autism symptoms and improve social and communicative skills in children (3-8 years old).

Strengths of the PicASSo project are as follows:

- The PicASSo project utilized a Randomized Controlled Trial (RCT) design, which is considered a robust method for assessing the effectiveness of interventions. The inclusion of three conditions allowed for comparison between therapy as usual, parent-mediated PRT without NAO robot, and parent-mediated PRT with NAO robot, enhancing the credibility of the findings [109].

- The use of NAO robot in PRT sessions created an engaging and interactive environment for the children. The robot's ability to respond and provide rewards when appropriate initiation was shown may have contributed to improved initiation skills in the children [61].

- The project's findings indicated that ASD symptoms decreased in the PRT with robot condition, suggesting that employing robots with the PRT approach might be effective in improving social and communicative skills in children with autism [20, 204].

- The related study comparing the "mechanical bodily appearance" and "humanized bodily appearance" of the robot revealed that the latter elicited higher interest, happiness, and affect scores in children. This finding highlights the importance of robot design and appearance in influencing emotional responses in children with autism during interactions [219].

Limitations and challenges of the PicASSo project:

- The PicASSo project was conducted in a specific setting and with specific age groups. The *generalizability* of the findings to other populations or settings should be considered cautiously.

- The use of NAO robots and related technology may involve significant costs, which could be a challenge in the wider implementation and scalability of the approach.

Overall, the PicASSo project demonstrated promising results in employing NAO robots with parent-mediated PRT to improve social and communicative skills in children with autism. The RCT design and engaging robot interactions contributed to the study's credibility and positive outcomes.

2.1.8 Robotica-Autismo Project

Robótica-Autismo[3] project conducted between 2009 and 2017 years by the University of Minho and the Portuguese Association of Parents and Friends of Mentally Retarded Citizens of Braga. The project's goal was to help children with ASD improve their emotion recognition and communication with others skills [51].

While the project demonstrated positive engagement and learning outcomes, a critical analysis highlights some strengths and limitations to consider.

Strengths of the Robotica-Autismo project are as follows:

- The project employed various robots, such as Lego Mindstorms NXT and KASPAR, to engage children with ASD and improve academic, cognitive, and socio-emotional skills [49, 50, 200]. The practical application of robots in therapy and intervention showcased their potential in supporting children's development.

- The implementation of algorithms for *detecting stereotyped behaviours*, such as hand-flapping, demonstrated the project's focus on understanding and addressing specific challenges faced by children with autism [175].

- Through interactions with the KASPAR robot, children with autism could develop body awareness and engage in appropriate tactile interactions [49]. This

[3]http://robotica-autismo.dei.uminho.pt/

highlights the potential of robots in assisting children with sensory processing challenges.

- The recent work using the ZECA (Zeno) robot to detect facial emotions and teach emotion recognition and imitation skills showed promise in addressing social communication deficits in children with ASD [201].

Limitations and challenges of the Robotica-Autismo project:

- Some studies within the project involved a relatively *small number of participants*, which might limit the generalizability of the findings [49, 50, 200].
- The project's duration spanned several years, but there were limited *long-term follow-up evaluations* to assess the sustainability of the improvements in children's skills.
- ASD is a spectrum with diverse characteristics and needs.
- Some studies did not have control groups, making it challenging to draw definitive conclusions about the specific impact of robot-assisted interventions compared to other approaches

Overall, the Robotica-Autismo project demonstrated that the robot successfully engaged children who have been diagnosed with autism. Therefore they had positive learning outcomes in developing socio-emotional skills [205].

2.1.9 SARACEN Project

The SARACEN (Socially Assistive Robots Autistic Children EducatioN) project (2013-2015) focused to provide tools for assistance throughout therapy and early diagnosis for children with ASD and designing interaction scenarios using the NAO robot [146].

While the project explored important aspects of robot-assisted therapy, a critical analysis highlights both its strengths and limitations. Strengths of the SARACEN project are as follows:

- The project focused on children diagnosed with high-functioning ASD, which allowed for targeted interventions and addressed specific challenges faced by this particular group of individuals [146].
- The structured therapy sessions with increasing difficulty levels (Easy, Medium, and Hard) provided a systematic approach to engage the children with autism and enhance their skills progressively.
- The development of robot algorithms for motion tracking, face scanning, child detection, attention measurement, voice analysis, and emotion recognition showcased advancements in technology to capture and respond to the child's behavioural responses effectively.
- The project reported improvements in attention and engagement of the children with autism during the therapy sessions, suggesting the potential of robot-assisted interventions to increase the child's involvement in therapeutic activities.

Limitations and challenges of the SARACEN project:

- The project involved only three children with high-functioning ASD, which limits the generalizability of the findings to a broader population [146].
- The average therapy session duration of 20 minutes might not be sufficient to fully assess the long-term impact of the robot-assisted interventions on the children's skills and behaviours.
- While supervised sessions allowed for controlled interactions, they might not fully reflect real-life scenarios where children interact with robots independently or with less supervision.
- The project's focus on the NAO robot and specific behavioural responses (attention, voice, facial expression) might not encompass the full range of social and communicative challenges faced by children with autism.

Overall, the SARACEN project made valuable contributions to the field of robot-assisted therapy. The focus on high-functioning ASD and the use of structured therapy sessions with individualized support demonstrated the potential of robots as effective tools to engage and support children with autism in therapeutic activities.

2.1.10 EMBOA Project

EMBOA[4] (Affective loop in Socially Assistive Robotics as an intervention tool for children with autism) project was supported by the EU Erasmus Plus Strategic Partnership for Higher Education Programme from 2019 to 2022.

While the project explored the potential of affective computing technologies in RAAT, a critical analysis highlights its strengths and challenges.

Strengths of the EMBOA project are as follows:

- The project's focus on combining emotion recognition technologies with social robots addresses the crucial aspect of emotional understanding and communication in children with autism [15].
- The project utilized audio and video recordings, physiological signals (E4 wristband), and eye gaze data (Gazepoint Eye Tracker) to create the EMBOA dataset, allowing for a comprehensive analysis of children's interactions with the social robot [2, 122, 139].
- The project's development of emotion recognition guidelines[5] for robot-assisted therapy provides valuable insights into effectively incorporating affective computing technologies in interventions for children with autism.
- The project's collaboration across four countries (North Macedonia, Poland, Turkey, and the United Kingdom) showcases the significance of international cooperation in advancing research and sharing resources [2].

[4] https://emboa.eu/

[5] $https://emboa.eu/wp-content/uploads/2022/10/AER-RIA\_v1.2.pdf$

- Making the EMBOA dataset[6] publicly available allows researchers worldwide to access and utilize the data, fostering further advancements in the field of social robotics and autism therapy.

Limitations and challenges of the EMBOA project:

- The project involved 29 children diagnosed with ASD, which might limit the generalizability of the findings to a broader population of children with autism [2].
- The number of therapy sessions attended by the children (2-12 sessions) might not provide sufficient insight into the long-term effects of robot-assisted interventions on social and communicative skills.
- The adoption of affective computing technologies and social robots in real-world settings might face challenges related to interoperability, scalability, and cost-effectiveness.
- Children with autism have diverse needs and preferences, which may require highly individualized robot interactions, making it challenging to design a one-size-fits-all approach.

Overall, the EMBOA project adds valuable insights to the growing body of research exploring the potential of social robots as intervention tools for children with ASD.

2.2 Social Robots for Children with ASD

Recent research indicates that children diagnosed with ASD readily accept social robots, which positively impact the development and enhancement of imitation abilities, joint attention, eye contact, behavioural responses, and repetitive and stereotyped behaviours [20, 151]. According to some reviews, children with ASD practice life skills more efficiently with robots than with people [17, 64, 151]. Humanoid robots

[6] $https://emboa.eu/wp-content/uploads/2022/10/EMBOA\_dataset.pdf$

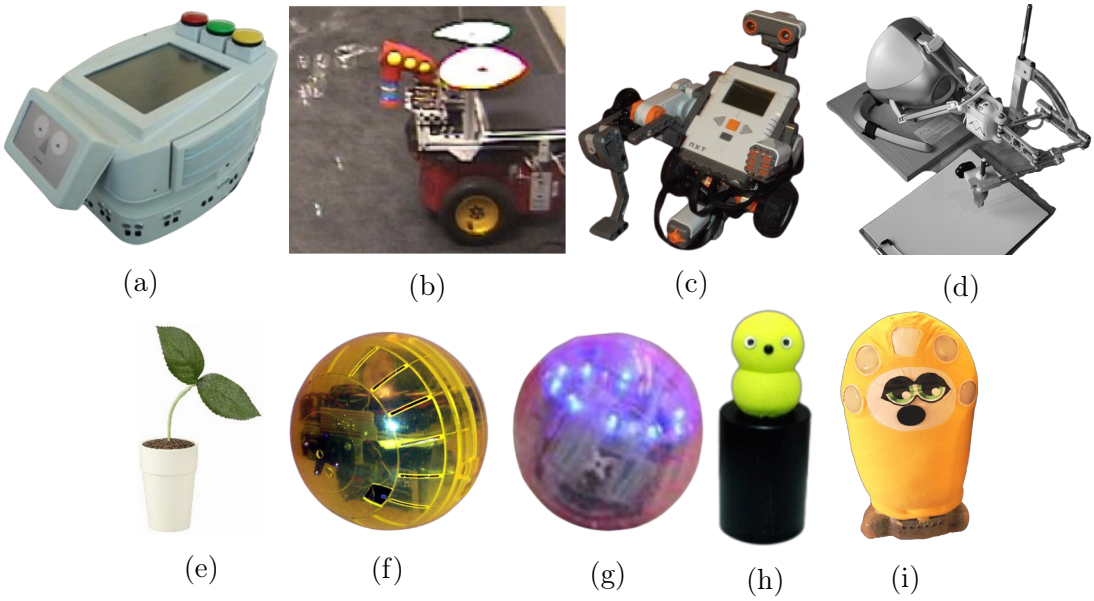

Figure 2-1: Functional robots: a) IROMEC [218], b) Bubble Blowing [71], c) Lego Mindstorms NXT [200], d) haptic device [147], e) Pekoppa [85], f) Roball [138], g) QueBall [183], h) Keepon [119], i) TeoG [29]

can be used in therapy to teach and practice social skills in a non-judgmental environment, which may help children with ASD feel less anxious and stressed. However, humanoid robots are not replacements for human therapists or social interaction [187]. They are best used as a supplement to therapy, and should always be used under the guidance of a trained therapist or medical professional.

A variety of robots are available in the research aimed to support and improve the social skills and education of children diagnosed with ASD. Social robots can be categorized in various ways, such as physical characteristics, the aim of use, and interaction capabilities. Most researchers classified social robots based on their morphological appearance into three groups: anthropomorphic, zoomorphic and functional robots [151, 187, 237]. However, Kouroupa et al. [116] categorized the social robots into four groups: humanoid, animaloid, and other that included a robotic arm and a plant robot. Moreover, all researchers stated that there were some robots that can be grouped as a union of two groups [151]. Bartl-Pokorny et al. [15] based on their morphological characteristics grouped social robots into five categories: mobile robots, animal, humanoid, ball-shaped robots, and others.

Functional Robots. Certain robots are not created to resemble any biological entities, including those that are humanoid or animal-like. These non-biomimetic robots possess varying appearances depending on their purpose, but they typically exhibit a simplistic design, user-friendly operation, and a toy-like aesthetic (Figure 2-1). They have different shapes and sizes, look like electrical devices and might not be covered with fabric.

Zoomorphic Robots. are also called "animal-like" robots (Figure 2-2). They are easy to interpret as they have an appropriate physical form. Most zoomorphic robots are able to express basic social cues. However, they are simpler than anthropomorphic robots. In the following paragraphs, some of the zoomorphic robots used in RAAT are described.

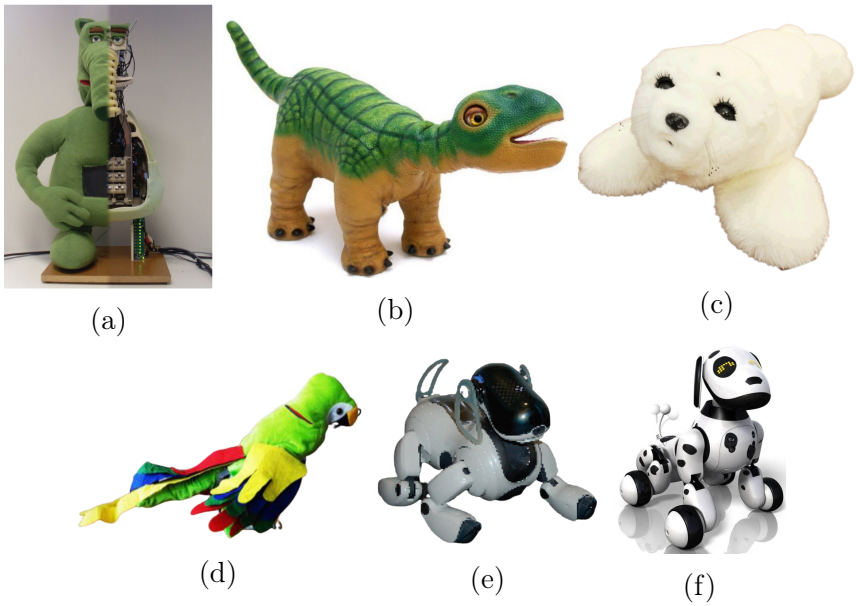

Figure 2-2: Zoomorphic robots: a) Probo [221], b) Pleo [106], c) Paro [165], d) KiliRo [21], e) Aibo [78], f) Zoomer [199]

Humanoid Robots. The humanoid robot features arms, legs and a head that mimic the human form (Figure 2-3). Some humanoid robots may possess facial attributes such as eyes, mouth, and nose, while others may only model the upper body starting from the waist. Equipping the robot with human-like facial characteristics (mouth, eyes, nose, etc.) is capable to facilitate reciprocal focus shared by the child with ASD and the robot. Humanoid robots can be programmed to show body language and facial expressions that are easily interpretable and comprehensible for children with ASD. They have shown potential in the therapy of ASD. According to studies, they can help children with ASD significantly improve their social skills, communication, joint attention, emotional regulation, and ability to recognize facial expressions.

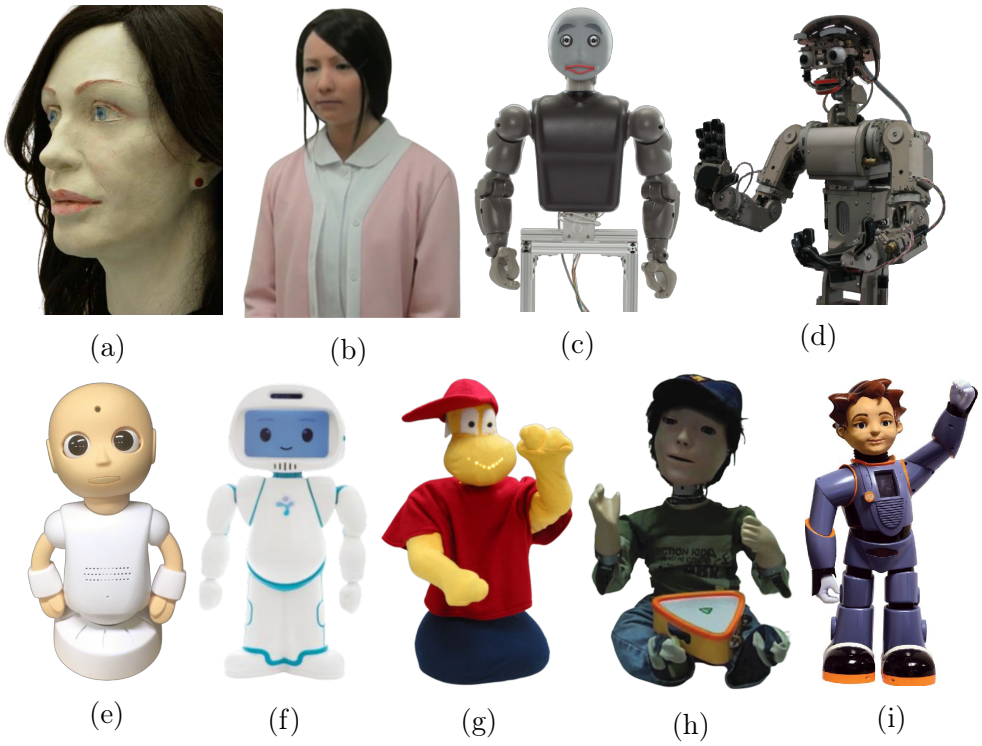

Figure 2-3: Humanoid robots: a) FACE [154], b) Actroid-F [120], c) Bandit [70], d) Infanoid [119], e) CommU [121], f) QTrobot [48], g) Tito [68], h) KASPAR [173], i) ZENO [62]

Overall, social robots may offer a potential pathway for improving the social and communication skills of children with ASD. Nonetheless, for social robots to be efficient in assisting children with ASD, they must be engaging. Engagement is fun-

damental to interaction because it helps keep individuals invested and interested in the interaction. For children with ASD, engagement can be particularly important because it can help them stay focused and motivated during therapy sessions.

2.3 Long-term HRI Studies for Children with Autism

Designing a long-term interaction between a human and a robot is complicated and requires significant time and resources from researchers in the field of HRI. Despite the challenges, longitudinal studies can provide valuable insights into the dynamics of HRI and how they change over time. In order to enable prolonged interaction, it is crucial for the robot to maintain its learning function and durability over time [197]. Such long-term interactions are crucial for advancing the field and harnessing the potential of robots in therapeutic and educational settings. Shibata [197] highlights the significance of maintaining a robot's learning function and durability over time to enable prolonged interaction, and this notion forms the cornerstone of our research.

Long-term interactions between humans and robots are not merely about extending the duration of engagement but also involves creating an environment that fosters a sense of continuity and predictability [141]. This sense of continuity is vital for building effective relationships between humans and robots, especially in therapeutic contexts. Baxter et al. [16] and other studies have demonstrated that establishing a long-term relationship with a companionable robot can have a positive impact on social bonding and overall well-being.

Weitlauf et al. [229] conducted a meta-analysis of traditional interaction-based therapy and behavioural interventions and found that the duration of therapy can vary from 3 to 16 weeks. Interestingly, their findings suggest that there is no clear correlation between the number of intervention hours and therapy outcomes. However, it is worth noting that earlier interventions tend to yield better results. In some cases, traditional therapies like Applied Behavior Analysis (ABA) and Intensive Behavioral Intervention (IBI) may extend beyond five years[7], highlighting the need for

[7]https://www.autismspeaks.org/applied-behavior-analysis

long-term therapeutic approaches.

A home environment is a common setting for long-term studies, as it provides a naturalistic and comfortable space for participants. Baraka et al. [13] identified the home as an example of this type of study. Pakkar et al. [145] and Clabaugh et al. [44] leveraged this setting for their studies, involving socially assistive robotic (SAR) systems for children with ASD. Their research underscores the potential for long-term engagement and positive interactions with robots in the familiar surroundings of the home.

Scassellati et al. [188] conducted a month-long study involving the Jibo robot, demonstrating the adaptability and positive impact of robots on children's emotional and social development when integrated into daily routines. This study provides further evidence of the feasibility and benefits of long-term interactions in home environments.

Francois et al. [78] explored the use of a dog-like Aibo robot in a UK school for non-directive play therapy sessions. Up to 10 sessions of the studies were conducted once every week. Their study highlights the importance of tailoring therapeutic interventions to each child's specific needs and capabilities. The results showed that each child demonstrated unique progress in one of the three aspects: Affect, Reasoning and Play. The achieved results emphasize the value of personalized, long-term approaches.

Robins et al. [170, 174] conducted longitudinal research involving humanoid robots and explored the effects of prolonged exposure on children's behaviours and engagement. Their findings suggest that open interactions with robots can lead to noticeable improvements in targeted behaviours, fostering positive outcomes over time.

Srinivasan et al.'s study [207] conducted a study that involved 36 children with ASD, aged 5 to 12 with rhythm interventions using Rovio and NAO robots. Over an eight-week period, spanning 32 sessions, these interventions were meticulously monitored to evaluate their effects on maladaptive and repetitive behaviours, and emotional states. The findings of the study showed a reduction in negative behaviours within the rhythm group and a consistent increase in interest observed in both the

robot and rhythm groups. These observations underscore the potential of movement- and music-based activities in enhancing engagement and social interaction among children with ASD.

In a parallel investigation, Huskens et al. [92] conducted a study in the Netherlands, targeting typically developing children and their siblings diagnosed with ASD. The study aimed to assess the impact of a robot-mediated intervention rooted in LEGO® Therapy. The results were striking, with improvements noted in collaborative play behaviours. Notably, children with ASD exhibited heightened interaction initiations, improved responsiveness to instructions, and increased engagement in joint actions with their siblings. These findings underscore the potential of robotics to facilitate cooperative play and social interaction among siblings with diverse needs.

Otterdijk et al. [144] carried out a study involving 25 children diagnosed with autism in the Netherlands to assess the long-term impact of CRI using PRT with a robot. The game-based PRT session lasted 15-20 minutes. Each child attended a total of 20 sessions. As part of more extensive research [20, 115] (Section 2.1.7), the investigators hypothesized that the children's attention and engagement would decrease over time, as prior studies indicated that children might lose interest and disengage. Contrary to expectations, the results revealed that the children's engagement and attention towards the robot and the game remained stable. Moreover, the children's awareness and involvement with their parents during therapy exhibited a linear increase over time; however, no such increase was observed concerning the therapist.

Kozima et al. [118] engaged in extensive field observations of preschoolers with and without developmental challenges over four years, documenting the interactions between children and the Keepon robot. Their research challenged prevailing notions about autism and highlighted the potential for robots to facilitate spontaneous exchanges of mental states.

In conclusion, these studies not only provide valuable insights into the design and performance of robotic technologies but also challenge our understanding of developmental differences and social interactions. Additionally, the findings challenge widely

held beliefs about autism and highlight the importance of understanding the needs and motivations of individuals with developmental differences. Overall, long-term research is essential for advancing the field of robotics and uncovering the full potential of HRI.

2.4 Personalization in RAAT

There have been several studies that aim to personalize robots to enhance their effectiveness in HRI. For instance, Andrist et al. [6] and Belpaeme et al. [19] investigated the impact of changing a robot's personality with different gaze behaviours or speech recognition on users' perceptions and the duration of their interaction with the robot. The findings of these studies were significant in demonstrating the potential benefits of designing methods for robots to adapt to their users' personalities.

The study by Sung et al. [209] utilized the Roomba robot to evaluate the use of a personalization toolkit for assessing progress over time by soliciting feedback. The results of this work also confirmed that children were more engaged with personalized robots. Moreover, some studies have focused on personalizing robots or therapy sessions based on children's interests, resulting in the development of a user model and the adjustment of CRI [90].

Kennedy et al. [105] provided a valuable perspective on the utilization of a robot that employed social behaviours, incorporating various gestures like personalization (e.g., greeting the child using their name) and eye contact (e.g., directing its view towards the touchscreen or the child). The robot adapted to a child's learning requirements to enhance performance. Complementing these previous works, Janssen et al. [98] highlighted the importance of personalizing robots to foster children's interest in interacting with them, as the social behaviour of the robot can be helpful in facilitating learning and motivation during appropriate occasions.

2.5 Engagement is Fundamental to Interaction

Engagement is a crucial social skill that may raise social robots' acceptance, effectiveness, and usability. It is a common and important term across many disciplines. Engagement is seen as a challenging term that accepts several definitions in the HRI literature [47, 86].

Perski et al. [152] conceptualized engagement in terms of individual experiences and behaviours targeting attention, interest, and affection. Based on their systematic review, they developed: two artificial notions:'Engagement as subjective experience' and 'engagement as behaviour'. During a short interaction with a system, engagement as a subjective experience manifests as a mental state characterized by focused attention, intrinsic interest and enjoyment, a balance between difficulty and competence, and temporal decoupling. However, engagement as behaviour is the level of commitment over a prolonged time, divided into active and passive. They identified two measures of engagement: "subjective" and "objective." Qualitative methods like the think-aloud technique or interviewing are examples of subjective measurements. The automatic tracking of usage patterns, such as the number of logins (time spent online and the quantity and kind of content used during the intervention period), cardiac activity, respiratory depth, electrodermal activity, and eye tracking can serve as an objective metric in contrast. Overall, Perski et al. [152] defined engagement as a dynamic process that varies both within and between people throughout time.

Engagement is a multidimensional construct that includes the following interdependent components [79]:

- Behavioral engagement, which defines engagement or active participation in educational activities.
- Emotional engagement, which defines a child's curiosity towards the tasks
- Cognitive engagement, which defines a child's willingness to learn new knowledge and abilities.

In many studies, engagement has been conceptualized from the perspectives of

technology, interaction, and education [239]. Education-wise, engagement is the main variable that predicts positive student outcomes [129] and provides educators with the possibility to alter their teaching strategies depending on learners' needs [239].

Kelders et al. [104] in their systematic scoping review proposed a more comprehensive understanding of engagement. They identified engagement as a multi-faceted construct with elements of behaviour, cognition, and emotion, yet more nuanced questions emerge when discussed separately. They categorized seven groups of engagement types: transdisciplinary, digital, work, societal, health, customer, and student engagement. Engagement is primarily viewed as a condition of being engaged with something in all groups. However, the fact that nearly all groups also consider engagement as a process. Sometimes it's thought that identifying the state of engagement itself is less significant than the process of being involved, sustaining engagement, disengaging, and re-engaging [104].

Researchers, therapists, and teachers commonly employ behavioural observations, self-assessment, and various technological tools to measure engagement levels in children with ASD [67, 103]. However, these methods are not always effective in measuring engagement levels for certain reasons. First, therapists may lack the expertise required to apply these methods accurately due to their complexity. Second, studies reviewed in Keen et al. [103] draw heavily on the effects of group size in the samples and may be inadequate for measuring engagement in children with ASD, who require a personalized approach. Anzalone et al. [7] and Salam et al. [180] have emphasized the importance of engagement as a key variable in designing intelligent systems that can adjust to the user's level.

2.6 The Components of Engagement

According to Corrigan et al. [46], engagement has two components:

- cognitive - characterized by attention and focus,
- affective - expressed by enjoyment.

The degree and quality of these elements encompass directedness, levels of attention, interest, engagement, and engagement quality. The term 'directedness' describes how a user's bodily components are momentarily aligned. Examples include how the head, eyes, torso, and movement routes are oriented. The level of attention is calculated by the duration of eye gaze on a particular object or place. The level of interest is closely related to the level of engagement and is based on the stored attention level. Additionally, they identified three different engagement levels [46]:

- uninterested in the action space or scene;
- superficially interested in the action or scene;
- engaged in the interaction.

O'Brien et al. [142] have identified several attributes of human engagement, such as affective appeal, sensory appeal, aesthetics, goal-directedness, attention, challenge, motivation, perceived control, feedback, and meaningfulness. These attributes indicate to what extent the interaction between an individual and a robot provides quality-oriented experiences. Thus, engagement can be treated as a quality indicator of HRI [198, 203]. This interaction can also be viewed from the theory of motivation, in which engagement is characterized as the motivational construct. When an individual feels enjoyment, puts effort, and shows commitment during a particular task, these states imply a higher level of engagement. In autism therapy, children are often exposed to a robot-mediated intervention to create a safer environment that allows them to retain immersive and engaging social communication. Engagement is critical for success when CRI is responsive and personalized to reach specific research goals.

As the field of HRI continues to develop, more researchers are studying engagement using different cues [143]. However, there is no single definition to identify engagement. All research works can be divided into two categories: those that consider engagement as the state of being engaged or disengaged [95], and those that take engagement as a process. For example, engagement is described by Sidner et al. [198] as "the process by which interactors start, maintain, and end their perceived connections to each other during an interaction."

Based on the interaction's receiver, user engagement may also be categorized. The user can interact with the agent, the work at hand, or the entire system (agent and task together), for instance, in human-agent interactions. The first scenario is known as social engagement, and the second is known as task engagement.

Researchers have noted that children with ASD often exhibit a low level of engagement, which can limit their ability to acquire and practice social skills in the outside world [111, 136]. Keen [103] has emphasized that environmental engagement is especially important for developing children's learning abilities, as it involves engaging with peers, parents, teachers, therapists, and other agents in their surroundings. Previous research has also established that high intensity of engagement plays a crucial role in developing children's learning abilities [136].

To summarize, engagement is fundamental to interaction because it is the measure of how invested or interested someone is in the interaction. It plays a crucial role in various fields, such as education, psychology, and HRI, as it can affect learning, motivation, and user experience. Recognizing engagement is important for various reasons, including improving learning outcomes, detecting mental health problems, and designing effective human-computer interfaces. Several models have been developed to recognize engagement, including physiological, behavioural, and self-report models. These models use various indicators, such as heart rate, facial expressions, and self-assessment, to measure engagement.

2.7 Models for Engagement Recognition

The concept of engagement is typically observed in various (non)verbal therapeutic tasks that involve interacting with robots. These tasks may include imitating the robot's behaviour, turn-taking, and making eye contact with the robot. According to studies by Lemaignan et al. [127] and Perski et al. [152], engagement signals are a sense of togetherness or with-me-ness that can be measured during social interaction and interactive tasks.

The recognition of engagement is crucial in advancing our understanding of com-

plex multimodal interactions. Anzalone et al. [7] proposed a methodology for evaluating engagement between a human and a robot in a triadic interaction. This methodology is based on dynamic social cues, such as joint attention, synchrony, and imitation, and static social cues, such as the focus of attention and head/body posture stability. The authors highlighted that their approach, which uses an RGB-D sensor and 2D histogram, is easy to use and does not require a marker-based system. This enables them to measure engagement in a natural environment without any restrictions or complexities.

In another study, Drejing et al. [67] proposed the use of the ADOS-G tool to evaluate engagement. The tool was evaluated by a therapist and involved activities such as imitation, turn-taking, and joint attention tasks. Participant ratings were recorded using a 5-Likert scale for the smiley face system, while observer ratings used an annotation program (ELAN) and a 5-Likert scale ranging from intense non-compliance (0) to intense engagement (5).

Research by Tapus et al. [213], used video data of the interactions between NAO robots and children with ASD to evaluate participants' engagement. They utilized Kinect sensor to track skeletons and three video cameras to capture the interactions. The experiment involved four children from Romania aged between 2.8 and 6 years old (M=4.2 years). During the dyadic and triadic interactions, the researchers measured five variables: the amount of spontaneous and prompt initiations, the length of eye contact and smiles, and the incidence of eye contact between two participants. The study found that when the NAO robot changed the colour of its eyes, and kept the children's attention. Additionally, it was shown that at the beginning of the interaction, the robot had a greater impact on the strength of good feelings. However, the quantitative outcomes were inconsistent, and there was considerable variability in responses to the NAO robot.

Poltorak et al. [155] employed Convolutional Neural Networks (CNNs) to evaluate the quality of HRI by analyzing images and calculating the emotion-dependent engagement metric, which could be categorized as engaged, neutral, or disengaged. The robot's learning algorithm required continuous adaptation and adjustments to

ensure dynamic and effective interaction. The authors used several metrics to evaluate engagement, including visual cues [7] such as tracking eye gaze, head direction, back posture, and facial expressions [69], as well as external sensors like pose estimation, pressure sensing [65], or body motion tracking [190]. Additionally, they utilized the concept of with-me-ness proposed by Lemaignan et al. [127] as another metric to assess engagement.

The proposed method for assessing engagement using CNN was tested on a small dataset of videos for seven emotions. However, these measures may not be entirely reliable as they can vary with a person's personality. To train the neural network, three datasets of grayscale 48x48 images were used, and the learning algorithm required time and task-varying changes to perform dynamic and effective interaction. The datasets used in this study comprised SFEW (Static Facial Expressions in the Wild) with 1,766 images, FER (Facial Expression Recognition) containing 35,887 images, and TFD (Toronto Face Database) which includes 4,178 images. The training process was prone to overfitting, which was reduced by adding a dropout layer to the neural network, which rejects specific neurons during learning. The learning rate was also changed from constant to annealing. The accuracy achieved by the model was 72.5%, which increased to 80% after optimising the results using the transfer learning technique. The model was trained on ImageNet-VGG-F (22 layers: 5CONV and 5FC) and retrained in the last layer using different techniques such as initialization, learning, and testing.

To detect user engagement automatically, q 12 [148] conducted an experiment in a real classroom environment in the UK with children aged 11-13 years old who played geography-based learning activities on a touch table assisted by the NAO robot. They used a Kinect sensor to extract lean body position, eye gaze, facial expressions, head direction, and affective state using electro-dermal recordings from a Q Sensor. The study's objective was to employ a set of short tests to identify engagement markers, including detecting lower and upper bounds of engagement within a task and fostering social connections between the participant and the robot.

Using low-level optical flow-based characteristics, Rajagopalan et al. [158] created

a way to assess a child's level of engagement. Based on the discovered underlying behavioural patterns of the child, they employed a two-stage underlying Conditional Random Field (HCRF) framework and Support Vector Machines (SVM) model to predict engagement. The Multimodal Dyadic Behaviour (MMDB) dataset, which comprises more than 160 Rapid-ABC sessions of adult-child social interaction, was used to assess the proposed model. According to a scale from 0 (easily engaged) to 2 (extremely difficult to engage), adults in this dataset judged how difficult it was to keep the children interested in them. According to the study, the two-stage (HCRF + SVM) method increased average engagement prediction accuracy by 3.3%, leading to a value of 79.7%. Moreover, it should be noted that the current method can extract low-level features such as keypoints more robustly than high-level features such as hand gestures and head poses.

Kapoor and Picard's proposed system [102] seeks to categorize children's interest levels in a learning task by integrating postural changes and facial expressions along with information regarding the task of the learner. They proposed a recognition system that categorizes emotional states connected to interest in children who are attempting to solve a computer problem using a mixture of Gaussian Processes. In order to do this, they gather nonverbal indications from a camera and a pressure-sensing chair, such as facial characteristics, behaviours, and postures. Through their proposed multisensor classification scheme, they achieved a recognition rate of 86%.

Tan et al. [211] suggested an automated engagement recognition system to monitor the actions and effects of children with ASD. The 12 actions from the NTU RGB+D dataset were used to assess their system [194]. To extract features, they utilized OpenPose, which extracted 25 body features from the short videos. They employed six standard machine learning algorithms, including Decision Tree, Logistic Regression, Ridge Regression, Random Forest, Gradient Descent, and SVM, for classification. Based on their classification results, Random Forest achieved the highest accuracy (99%), while Ridge Regression achieved the lowest accuracy (2.6%). However, it should be noted that the evaluation was conducted on a limited set of actions and not on a specific engagement task or scenario. Therefore, the general use of

the suggested strategy in additional cases and tasks requires further evaluation and testing.

Alcorn et al. [3] investigated the perspectives of educators on employing NAO as a tool for education for pupils diagnosed with autism. The study included 31 participants from special schools in England who highlighted the significance of engagement and motivation in learning for children with autism. The educators highlighted that for these children, it is crucial to feel comfortable, safe, and secure, and the predictability and consistency of the robot as compared to humans can be advantageous. The visual consistency of the robot may also help children focus their attention on learning. However, the educators emphasized that robots are not toys, and understanding the differences between humans and robots depends on the cognitive ability and child's age. The educators suggested that the robot should be personalized for each child and adapted to their individual needs, but they also highlighted a potential drawback that children with autism might not learn how to understand and communicate with other people if they only interact with robots.

Hence, future work should incorporate stakeholders in the planning and execution phases, such as educators, parents, and children with autism, to guarantee that the efforts have a lasting and direct influence on those who require it the most [192]. The phase should start with a thorough joint examination of the presumed and anticipated advantages of robots for children with autism while weighing these against developmental, potential interpersonal, and resource-related costs.

In a multimodal setting, where data is generated from multiple input modalities with different representations and structures, learning can be challenging due to noisy and missing values [176]. To address this, the labelling approach can be applied. For instance, Srivastava et al. [208] proposed an MDBM (Multimodal Deep Boltzmann Machine) model that is composed of unimodal undirected pathways and put layers of hidden units between modalities to provide scaffolding for the differences in their statistical properties. Similarly, Rudovic et al. [178] proposed a multimodal active learning approach to optimize data selection and personalize the target classifiers for estimating engagement levels in children with autism during therapy. This involves

two sequential processes: training classifiers using majority voting and learning for active data selection using the reinforcement learning model's Q-function. The proposed approaches can improve engagement estimation and personalize the learning process for children with autism. However, more research is required to investigate their effectiveness in larger datasets and in real-world settings.

Considering the classification models, the trained multimodal engagement classifiers may be suboptimal because of the notably different ways of engagement appearances (head movements, facial expressions, body gestures, and positions) and their dynamics across children diagnosed with autism. In addition to the above-mentioned technical methods of engagement assessment of children during their interaction with a robot, it is also crucial to point out the impact of personalization of the procedures and therapies to suit the diverse needs of children with ASD and the typically developed ones. However, the achievement of personalization seems to face challenges, and lack consensus on researchers' parts [235].

The purpose of the current study is to develop an automatic recognition model of engagement for children with ASD that is based on data collected from personalized long-term robot-assisted therapy (RAT) in clinical settings with a humanoid robot NAO. This approach differs from previous studies (Jain et al. [97] and Rudovic et al. [176]) in its focus on long-term engagement assessment in clinical settings and its integration into both educational and clinical settings for children with autism. The study combines binary classification with multi-class classification problems, in order to capture the diverse ways in which engagement manifests in different individuals with ASD. The aim is to implement a model that can recognize user engagement in a personalized and adaptive manner, considering the diverse requirements of children with ASD.

Summing up, the studies mentioned above can be divided into two groups: detection of current engagement signs and interaction events [41]; and prediction using supervised models trained with data, such as task-based interaction, social, or physiological features [185]. The existing studies inform the experts, therapists and educators about the feasibility of applying automatic methods of engagement recognition

and offer advancements in the therapy of ASD in clinical settings. Yet, research still faces methodological constraints in automatic data-driven engagement recognition.

2.8 HRI Datasets

While interest in social robots is rising that can learn social interaction, the lack of publicly available HRI datasets is a major obstacle to progress in this field. Due to privacy issues, most HRI datasets are not accessible to the general public, especially for CRI. However, during our literature survey, we found several HRI datasets that have been made available to the research community. Researchers can use these datasets to train social robots to better interact with humans and improve the overall user experience. However, it is crucial to handle these datasets with care and respect for the privacy of the participants involved. Below we are providing a brief description of the HRI datasets.

2.8.1 Keepon Pro-Active Dataset

The "Keepon Pro-Active" dataset was collected by a research team led by Dr. Hideki Kozima from the National Institute of Information and Communications Technology (NICT) in Japan [117]. The dataset comprises over 400 hours of video recordings of interactions between preschoolers and Keepon. In particular, the researchers focused on studying the interactions between Keepon and children with ASD, and those with typical development. The dataset includes recordings of over 30 children (ages 2-4) at a daycare facility over the course of 100 sessions, or 700 kid sessions overall, for observations pertaining to autism [117, 118, 119]. The dataset includes video recordings from the robot's cameras, as well as audio recordings of the interactions. The researchers used these recordings to study how the children interacted with Keepon, and how Keepon's behaviour influenced the behaviour and emotions of the children. The dataset is available for research purposes upon request. It has been used in a number of studies exploring the use of robots for therapy and education for children with ASD.

2.8.2 DREAM Dataset

The DREAM dataset was introduced by Billing et al. [23]. It consists of more than 300 hours of video recordings with 61 children with autism (9 female, aged 3-6 years) during two therapy conditions: with a robot (RET, n = 30) and with humans (SHT, n = 31). Every therapy session was recorded via the same sensory therapy table. Three high-resolution RGB cameras and two Kinect (RGB-D) cameras were installed on this table. These cameras, along with sensor interpretation methods, were used to gather data about the child's verbal statements, movements, eye gaze, facial expressions, and position.

Every intervention was split into 3-6 sections, adhering to a task script that outlined the task and directions provided to the child. A total of eight interventions were recorded using two depth (Kinect) and three RGB cameras. This dataset was collected during the DREAM project (see Section 2.1.5). However, only a subset of all features was released publicly, including participants' age, gender, autism diagnosis (ADOS-G scores), therapy conditions (RET or SHT), date and time of recordings, therapy task (Turn-taking, Joint attention, and Imitation), 3D recordings of head position and orientation, body motion, and eye gaze characteristics. The authors used the JSON file format for data representation, and a Cartesian coordinate system was utilized to represent attributes of eye gaze, head gaze, and skeleton. Notably, the dataset does not include facial expressions or direct measurements of the child's performance during therapy.

2.8.3 MDCA Dataset

Rudovic et al. [179] proposed the Multimodal Dataset of Children with Autism (MDCA), which includes data from 35 children with autism (30 males and 5 females, aged 3-13 years) from two cultures: Asian (17 Japanese); and Europian (19 Serbian). The severity of the children's behaviours was assessed by utilizing the Childhood Autism Rating Scale (CARS). The children participated in only one short session, lasting an average of 25 minutes, which aimed to study the relationship between the

children's task-oriented engagement and valence and arousal of affective engagement was examined. The therapy was centred on a teaching method for emotion recognition and expression by associating images of expressive human faces with NAO's behaviour.

A high-resolution camera with a microphone was used to record the interaction. The engagement episode, from the initiation of a task until the subsequent engagement level was achieved, was rated using a 0-5 Likert scale. Moreover, valence and arousal were assessed using a 5-point ordinal scale, while the children's facial expression was measured using a 0-5 Likert scale. During coding, two coders achieved an agreement of 92.4% for estimating engagement, 75.8% for valence, 67.4% for arousal, and 69.8% for face expressivity.

2.8.4 MHHRI Dataset

The Multimodal Human-Human-Robot-Interactions (MHHRI) dataset [38] was collected in a controlled study examining interactions between two people and a robot. The participants asked each other personal questions while being recorded using two dynamic cameras mounted on the participants' heads, two biosensors and two static Microsoft Kinect depth sensors. The physiological signals (electrodermal activity (EDA), skin temperature, and 3-axis acceleration of the wrist) were recorded using the Q Sensor by Affectiva.

The dataset consists of about 6 hours' worth of multimodal recordings from 12 interaction sessions (7 sessions with an extroverted robot and 5 sessions with an introverted robot) with 18 participants, predominantly graduate students and researchers, including 9 females. Every session was approximately between 10 and 15 minutes.

The dataset also offers labels for personal characteristics and self-reported partner engagement, which were obtained using two different types of surveys. Celiktutan et al. [38] discovered that employing just first-person vision head motion signatures (FPV-HMS) produced the greatest results for engagement (mean F-Score = 0.59) and that individual characteristics improved performance for interaction.

2.8.5 PInSoRo Dataset

The PInSoRo dataset was collected to study under-specified free-play interactions between children and robots, with the aim of capturing behavioural tendencies in everyday interactions. The dataset includes 45 hours of hand-coded interaction data involving 45 child-child (42 female and 49 male) and 30 child-robot pairs (12 female and 18 male) [128].

The dataset includes entire audio files, thoroughly calibrated video frames, structural characteristics, 3D records of the face, game engagements, and annotations of social constructs. The data were collected using two short-range Intel RealSense SR300 RGB-D cameras to record the children's faces and audio, and a Microsoft Kinect RGB camera to record the whole interaction setting. The dataset employed a coding scheme that encompassed the level of task engagement (no play, adult-seeking, goal-oriented, and aimless), the level of social engagement (solitary, associative, onlooker, cooperative, and parallel), and the social attitude (pro-social, adversarial, assertive, frustrated, passive, or bored). Audio features were extracted utilizing the OpenSMILE toolkit with 33 ms-wide time windows, while the OpenPose library was employed to obtain 70 facial landmarks, the 18 points of the upper-body skeleton, and hand skeletons for each child. The dataset also contains in total of 17 action units along with their respective confidence levels, which were got by utilizing the OpenFace library [128].

By utilising the free-play sandbox task, the PInSoRo dataset attempted to provide the following experimental baselines: an 'asocial' baseline for the condition where children interacted with non-social robots and a 'human-social interactions' baseline for the condition where children interacted with other children.

2.9 Concluding Remarks

Current research on using robots to intervene in ASD frequently lacks methodological rigour, generalization of therapy [151], and insufficient understanding of how robots can enhance established interventions [64]. Moreover, RAAT has a significant dif-

ference in the number of sessions with relatively fewer long-term studies. There are several studies have investigated the employment of social robots in various settings, such as homes, educational environments, and hospital facilities, but there is a lack of a universally accepted definition and measurement framework for engagement.

Studies have shown that engagement and positive affect can be generated when children with autism interact with robots and such behaviour can even be generalized to co-present humans [72, 107, 112, 152, 168, 188]. Remarkably, SARs can engage children with ASD and hold their attention for longer than complex human-to-human interactions. The range of tasks and behaviours, like attention, interest, and motivation, forms the foundation of engagement, and the key to promoting effective learning progress lies in recognizing a child's emotional state during an interaction.

Recognizing engagement in children with ASD is challenging because they may not exhibit typical social cues. However, recent studies have demonstrated that social robots can measure engagement in children with ASD by monitoring their gaze and facial expressions and adjusting their behaviour accordingly. SAR, therefore, have the potential to be a valuable tool for helping children with ASD develop social skills, but they must be engaging. Recognizing engagement is fundamental to creating effective interactions that can improve learning, motivation, and overall well-being.

Furthermore, HRI datasets demonstrate the prospective utility of leveraging user interaction data to educate robots on social behaviours and assess the efficacy of robots in human interaction scenarios. However, the lack of publicly available datasets for CRI highlights the need for more efforts to create and share such datasets while ensuring privacy and ethical concerns. The availability of such datasets aids in the development of more effective interventions for children with ASD, utilizing social robots as a therapeutic mediator, thus heralding advancements in the broader domain of social robotics.

Chapter 3

Multi-Purposeful Robot Activities for RAAT

This chapter outlines the capabilities of the NAO robot and details 9 activity blocks ("Dances," "Songs," "Action Song," "Storytelling," "Follow Me," "Touch Me," "Imitations," "Social Acts," and "Emotions") suitable for various forms of autism. The goal is to offer individualized RAT that is tailored to the therapist's knowledge and judgment, ensuring that each child has a personalized and suitable experience.

3.1 Humanoid NAO Robot

The Aldebaran Robotics[1] manufactured a medium child-sized humanoid robot known as NAO for autism therapy [8, 196, 212, 216, 221]. NAO stands at a height of 58 cm and weighs 5 kg. The robot is made of plastic and contains a set of sensors and features, including a sonar range finder, one inertial board, two cameras, two IR emitters and receivers, two high-fidelity speakers, four directional microphones, eight pressure sensors, nine tactile sensors (Figure 3-1) [206]. With 14 DOF in its upper parts and 11 DoF in its lower limbs, NAO has a total of 25 DOF, providing it with great mobility. NAO's design and features make it an effective tool for individualized and engaging autism therapy.

[1]https://www.aldebaran.com/en

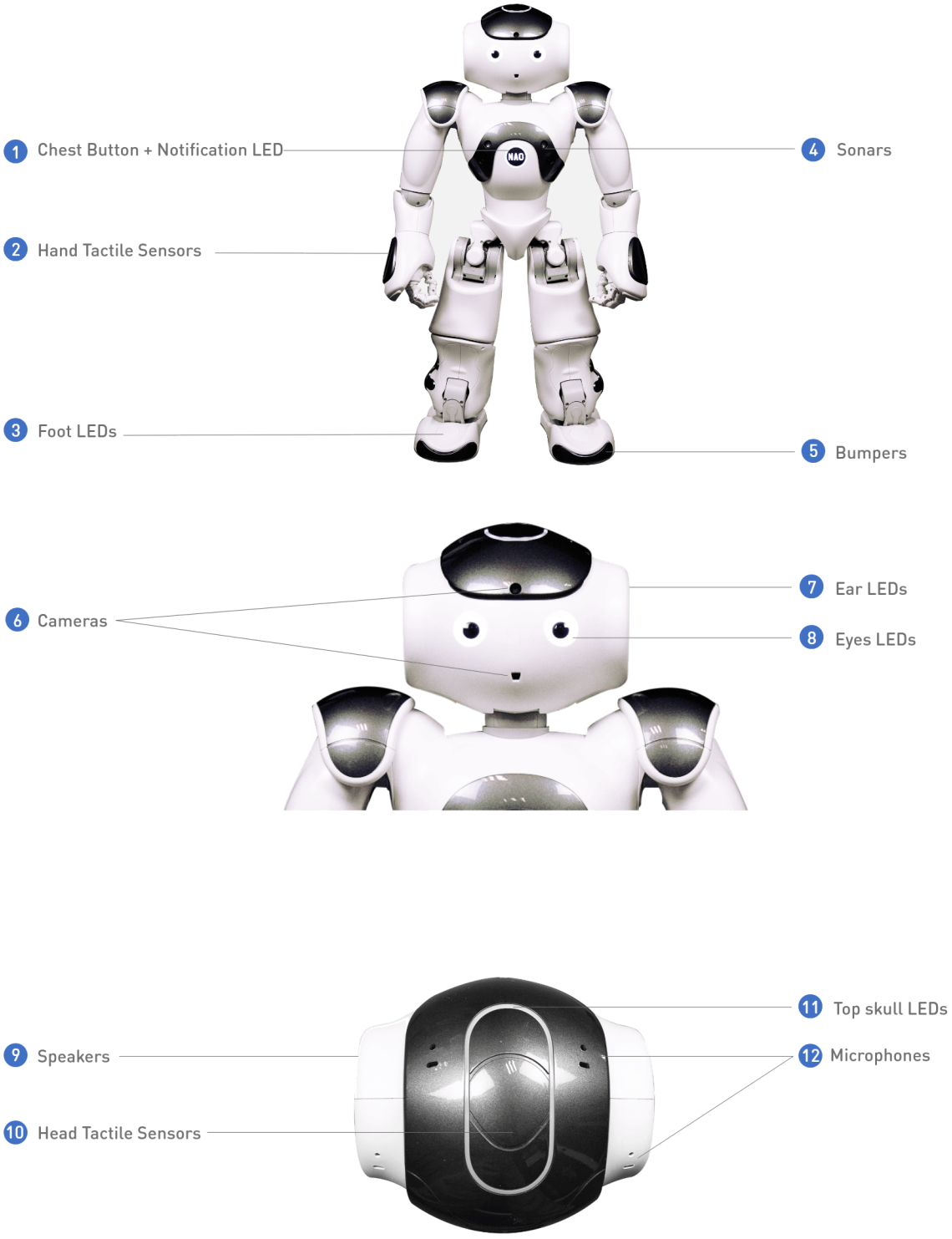

Figure 3-1: Overview of the NAO robot [206]

The review by Amirova et al. [5] reports that NAO was mostly used in RAAT in 33 studies published between 2009 and 2019 worldwide. NAO has been used for both diagnostic and therapeutic purposes in children diagnosed with ASD. Based on the review, NAO robots are mostly used for gesture and emotion recognition, imitation, display of emotions and performing verbal and nonverbal communications during RAT interventions [5, 58, 196]. Therefore, the NAO robot helps children with ASD to develop skills of social communication, understand facial expressions and emotions, retain eye gaze, and improve joint attention [7]. Children with ASD benefit from engagement with the robot during communication because of its ability to attract attention [34]. The humanoid social robot makes two-way communication easier, which may subsequently affect how well children interact. For example, in the study by Fuglerud et al. [81] the NAO robot is used as a language teacher in autism therapy. It should be noted that the quality of interaction assesses using timed feedback during HRI. Feedback from NAO robots typically takes the form of nodding, hand gestures (similar to a high five), and attention-grabbing spoken answers. NAO serves as a mediator for children who are diagnosed with autism to foster the growth of interpersonal and communicative abilities [92, 195, 196]. Saleh et al. [181] showed that the NAO robot provides therapeutic benefits including eye contact, engagement, visual attention, augmented social interaction skills, and positive expressiveness. Although several studies have noted that in addition to beneficial features, NAO has also some limitations. For instance, NAO moves slower than children, cannot tilt its head, makes a lot of noise, its eye gaze is unnatural, and it lacks the same social-emotional signals similar to humans. Therefore, it might be suggested to speed up the robot's movement and perform more realistic and natural eye contact for future research. In addition, researchers used the NAO robot because of its ability to use sensors and microphones to collect data [73] and actuators to implement different behaviours [184].

NAO features fundamental components that allow it to operate more naturally, including touch sensors, built-in facial and speech recognition, a text-to-speech engine, and a display of body and gestures posture. It also can communicate nonverbally

via its motions and LCD eyes. It can be programmed to be autonomous, semi-autonomous, or controlled by a human operator (WoZ). Researchers have found the NAO robot to be easy to operate [58], and it can detect and recognize pre-learned objects and faces. As a result, one of the most popular robotic platforms for autism therapy is the NAO robot [15, 151, 187]. In our own RAAT study for children diagnosed with autism, we also used the NAO robot.

3.2 ABA Principles

This thesis deployed a total of 24 robot activities that are based on Applied Behavior Analysis (ABA) principles, which include positive reinforcement, Picture Exchange Communication System (PECS), errorless teaching, and peer-mediated social skills training [123].

3.2.1 Positive Reinforcement

Positive reinforcement is an essential component of many behaviour-based therapies, including ABA. In our study, we used positive reinforcement almost in all activity blocks to promote positive behaviours in children diagnosed with autism. Positive reinforcement is based on the principle that behaviours that are followed by a positive outcome are more likely to be repeated in the future. Therefore, by praising children with rewards, we aimed to increase the likelihood that they would repeat the desired behaviours during the study. "Well done," "Keep up the good job," and "Perfect" are all examples of verbal praise. Nonverbal stimuli such as smiling, cheering, raising arms and clapping hands were also used.

In addition to verbal praise and nonverbal stimuli, we also used tangible rewards such as stickers. Stickers are a simple and effective way to reward children for their positive behaviour. When the sessions ended, by offering a choice of stickers, we allowed children to select the one that they preferred, which increased their motivation to perform well.

3.2.2 Picture Exchange Communication Systems

Picture Exchange Communication System (PECS) is an evidence-based communication system which can be effective in promoting communication in children with ASD. The system is based on the idea that children with communication difficulties may find it easier to understand and use visual symbols instead of spoken language.

The use of visual symbols in PECS allows children to express their needs, wants, and emotions more easily and also helps to develop their social skills. By using pictures to represent different emotions, children can learn to identify and label their own emotions and the emotions of others, which is an important step in developing emotional awareness and empathy.

In our study, we used PECS to encourage children to express emotions ("Emotions") and imitate ("Imitations") the sound of gestures of the robot. By showing the children images of emotions and asking them to match them with the appropriate gesture or sound, we provided a fun and engaging way for them to learn. Also in the "Storytelling" activity block, we used pictures to show the sequence of characters in the "The Bun" tale.

By incorporating PECS into our interventions, we provided children with a structured and visual way to learn, which can be especially beneficial for children with ASD who often have difficulty with abstract concepts and language. The use of visual symbols in PECS also provides a way for children to communicate independently, which can be empowering and boost their self-esteem.

3.2.3 Errorless Teaching

Errorless teaching is an approach used in ABA therapy to minimize errors and maximize learning. This technique involves providing prompts and cues to ensure that the learner completes the task successfully, without making any mistakes. The goal is to provide a setting where the learner feels supported and confident, which ultimately leads to increased success and positive outcomes.

In the study mentioned, errorless teaching was utilized for children to complete the

tasks successfully. By providing prompts and cues, the researchers aimed to prevent errors from occurring and increase the children's confidence and success. For example, when the child does not respond to the robot, a human therapist demonstrated the way of touching the sensors on the robot's arms and legs during the "Touch Me" activity block. Parents also gave their children prompts to help them perform proper behaviours.

Overall, the prompts and cues used in errorless teaching may include verbal prompts, gestural prompts, physical prompts, or visual prompts, depending on the individual's needs and abilities. The prompts are gradually faded over time as the learner becomes more proficient, ultimately leading to independent performance of the task.

3.2.4 Peer-mediated Techniques

In traditional ABA therapy, peers are often used to help teach new skills and promote social interaction. However, in our study, we used a child-like NAO robot as a substitute for human peers. The robot gave the children a fun and interesting opportunity to practice social interactions in a safe and supportive environment. By using the robot as a model, we aimed to increase the children's engagement and motivation to learn social and communicative behaviours. Throughout the intervention, we incorporated a peer-mediated technique in which the child participants observed the robot's behaviour and subsequently imitated its actions, focusing on social acts such as high-fives and hugs. The robot's behaviour was predictable and consistent, providing a low-anxiety and non-judgmental learning experience. By utilizing this technique, we aimed to promote socialization and improve social skills in children with autism.

Overall, the combination of positive reinforcement, PECS, errorless teaching, and peer-mediated techniques provided a comprehensive approach to support the child's behaviour during the study. By utilizing a range of tools and strategies, we aimed to maximize the children's engagement, motivation, and success in learning new skills.

3.3 Objectives of Activities

The International Classification of Functioning, Disability and Health for Children and Youth (ICF-CY) has provided a thorough list of therapeutic and educational goals [232]. In our study, we have implemented each activity with specific objectives based on the ICF-CY classification, as shown in Figure 3-2. These objectives include:

- emotional and social development,
- interaction and communication development,
- sensory development,
- motor development,
- cognitive development.

ICF-CY classifications	Dances	Songs	Action Songs	Story-telling	Follow Me	Touch Me	Imitations	Social Acts	Emotions
Sensory development									
Perceptual functions	✓	✓	✓	✓	✓	✓	✓	✓	✓
Communication and interaction									
Voice and speech functions		✓	✓				✓	✓	✓
Communicating - Producing - Nonverbal	✓	✓	✓	✓	✓	✓	✓	✓	✓
Basic interpersonal interaction	✓	✓	✓	✓	✓	✓	✓	✓	✓
Particular interpersonal relationships			✓	✓		✓	✓	✓	✓
Cognitive development									
Global intellectual functions	✓	✓	✓	✓	✓	✓	✓	✓	✓
Memory functions	✓	✓	✓	✓		✓	✓	✓	✓
Higher-level cognitive functions						✓		✓	✓
Thinking						✓			
Making decisions						✓			
Copying	✓	✓	✓				✓	✓	✓
Attention	✓	✓	✓	✓	✓	✓	✓	✓	✓
Energy and drive functions		✓	✓	✓	✓	✓			✓
Learning through action with objects				✓	✓	✓	✓	✓	✓
Undertaking tasks			✓		✓	✓	✓	✓	✓
Social and emotional development									
Emotional functions	✓	✓	✓	✓	✓	✓	✓	✓	✓
Experience of self and others			✓			✓	✓	✓	✓
Engagement in play	✓	✓	✓	✓	✓	✓	✓	✓	✓
Community, social and civic life			✓		✓	✓	✓	✓	✓
Motor development									
Body (mobility)	✓	✓	✓		✓	✓	✓	✓	✓
Objects (mobility)					✓	✓	✓		✓
Fine hand use (mobility)	✓	✓	✓		✓	✓	✓	✓	✓
Neuromusculo-skelatal functions	✓	✓	✓		✓		✓	✓	✓
Psychomotor functions	✓	✓	✓		✓	✓	✓	✓	✓

Figure 3-2: Objectives of the activities

3.3.1 Emotional and Social Development

Children diagnosed with ASD often struggle with social and emotional intelligence, making social and emotional learning (SEL) a crucial aspect of their therapy [227]. SEL can reinforce emotional self-awareness and empathy in children with ASD, who learn to express emotions and read others' facial expressions from infancy and continue to develop affective skills as they grow. Affective technologies, such as social robots, have shown promise in fostering social-emotional understanding and eliciting different emotional responses in children with ASD [76]. The following functions can be categorized as SEL:

- Emotional functions relate to the feeling and affective components of the mind's processes, including recognizing and expressing emotions.
- Experience of self and others involves developing awareness of one's own identity, body, position in the environment, and time, as well as understanding others' perspectives.
- Engagement in play includes purposeful and sustained engagement in activities with robots, either by oneself or with others.
- Community, social, and civic life involves engaging in aspects of community and social life, such as participating in group activities and demonstrating appropriate social behaviours.

All robot activities are designed to develop and improve emotional functions, promote engagement in activities, and facilitate social interactions by performing different types of play scenarios. For example, the "Emotions" and "Social Acts" activities aim to improve emotional recognition and empathy by asking children to imitate the robot's displayed emotions and movements. Meanwhile, the "Follow Me," "Touch Me" and "Imitations" activity encourages sustained engagement and social interaction by involving cooperative play between the child and the robot.

3.3.2 Interaction and Communication

Among all the symptoms of autism, the lack of communication and interpersonal skills is the most evident. Children with ASD often struggle to navigate social relationships and frequently rely on family members, such as mothers or siblings, for social interaction. Communication-driven activities are introduced to support the development of verbal language and nonverbal cues to recognize and convey social cues, thereby avoiding long-term and devastating effects. Peer-mediated interventions (PMI) are common mechanisms used to support social skills development, reciprocal exchanges, and relationships with peers [100].

Communication and interaction development include various functions, such as voice and speech functions, which refer to the production of sounds and verbal communication, including monosyllabic speech and non-speech sounds. Other functions include communicating through gestures, symbols, and drawings to express thoughts and needs, basic interpersonal interactions such as taking initiative, proximity, eye gaze, and responding to others, and specific interpersonal relationships such as collaboration, awareness, empathy, and patience.

In all activities, the robot serves as a peer to motivate children to be more active and social and to produce speech and non-speech sounds (see Figure 3-2). For example, during the "Songs" and "Storytelling" activities, the child might repeat words that the robot speaks, while during the "Imitations," "Emotions," and "Social Acts" activities, non-speech sounds are presented for the child to repeat as well.

3.3.3 Sensory Development

Sensory development is crucial for children's growth and can affect their perceptual functions, which include an understanding and interpretation of sensory inputs like auditory, visual, olfactory, gustatory, visuospatial and tactile perceptions [232]. However, perceptual functions do not include other functions such as consciousness, orientation, attention, memory, mental functions of language, seeing, hearing, and vestibular functions.

Research has shown that individuals with ASD often experience sensory abnormalities that regardless of age or skill level, should be wide-ranging and multimodal [124]. As sensory deficits can precede and predict social and communication barriers in childhood [166], autism-focused activities aim to support sensory stimuli and processing in various ways. Therefore, each robot activity in our study was designed to improve sensory stimulation (Figure 3-2). For example, all activities aim to improve auditory function by discriminating sounds, tones, and pitches. Additionally, the "Dances," "Action Song," 'Storytelling," and "Social Acts" activities improve visuospatial perceptions by distinguishing the relative positions of robots in their surroundings or with respect to oneself. The "Songs," "Follow Me," and "Touch Me" activities aim to improve tactile stimuli by encouraging children to touch the robot, while the "Imitations" and "Emotions" activities are designed to develop visual stimuli by involving pictures.

3.3.4 Motor Development

Motor difficulties are a challenge for children with ASD, affecting their physical development. According to recent estimates, 87% of children with ASD experience some form of motor difficulties, such as unusual movement, and difficulties with object manipulation and handwriting [22]. Traditional autism therapy typically includes occupational and physical therapeutic support, such as movement therapies incorporating different music types. The following functions are included in motor development:

- Body mobility involves getting into and out of different body positions and moving from one point to another, such as sitting, standing, squatting, or kneeling.
- Object mobility involves lifting and transporting objects from one place to another.
- Fine hand use involves coordinating actions to handle, pick up, manipulate, and release objects using one's hands, fingers, and thumbs.
- Neuromusculoskeletal functions refer to the range and ease of joint movement,

including the vertebral, elbow, ankle, wrist, shoulder, hip, knee, and small joints of the hands and feet.

- Psychomotor functions involve control over both motor and psychological events at the body level.

Music and dance-based activities have been considered and designed to address social and motor skill difficulties. For example, "Imitations," "Emotions," and "Social Acts" activities encourage a child to imitate robot movements and voice, while "Follow Me" improves walking, joint attention, and neuromusculoskeletal functions. Moreover, during "Imitations" and "Emotions" activities, children learn to hold and move objects from one place to another.

3.3.5 Cognitive Development

Joseph et al. [99] demonstrate that there are variations in the intellectual abilities of children with ASD, ranging from mild to severe mental impairments. Cognitive development is a complex aspect of autism therapy because children may not develop study skills due to varying levels of their autistic symptoms and learning conditions. The following functions are classified as cognitive development:

- Global intellectual functions, which include clarity, continuity, awareness and alertness of the wakeful state.
- Memory functions, which involve registering, storing, and retrieving information as needed.
- Higher-level cognitive functions include complex, goal-directed behaviours.
- Thinking, which involves formulating and manipulating, goal-oriented or not concepts and images.
- Decision-making, which involves making a choice, implementing it, and evaluating the effects of the choice.

- Imitation, which involves mimicking or copying the letters of an alphabet, sounds, gestures or facial expressions.
- Attention, which involves concentrating for the needed amount of time on an external or internal case.
- Energy and drive functions, which are the psychological and physiological mechanisms that cause one to persistently work toward satisfying requirements.
- Learning through action with objects, which involves symbolic and pretend play.
- Activity completion, which involves performing straightforward or complex, coordinated actions relating to the conceptual and practical aspects of a single activity.

In general, cognitive development involves gradually incorporating lower-order (e.g., memorization) and higher-order thinking (e.g., understanding) skills through mindful, thought-based, and multi-dimensional tasks. For example, higher-level cognitive functions are developed in activities such as "Emotions" (when the robot asks, a child looks at pointed images and replicates shown emotions) "Social Acts" (where a child repeats actions and sounds when the robot asks), and "Touch Me" (where a child should press the required sensor). These activities include complex, goal-directed behavioural functions. Additionally, the "Touch Me" activity aims to improve decision-making by choosing which robot's sensor to press and evaluating the effects of the made choice. When a correct sensor is pressed, the robot praises and applauds the child, otherwise, it remains quiet.

3.4 Activity Blocks

All robot activities are grouped into nine blocks. Two activities were downloaded from GitHub and the "Thai Chi" dance activity behaviour was downloaded from the Aldebaran Store. All of these activities were adapted to local languages. The remaining behaviours were designed and implemented at the HRI Lab at Nazarbayev

University (NU) using an iterative interaction design process [184]. The video of the activity demonstrations can be seen at this link: bit.ly/rat-nu.

3.4.1 "Dances" Activity Block

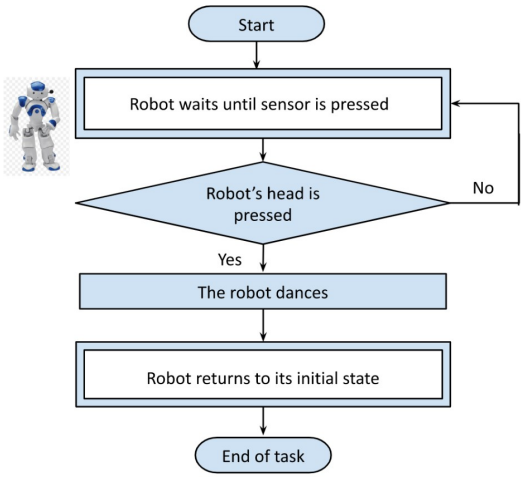

Figure 3-3: Activity diagram for "Dances"

The "Dances" activity block aims to promote emotional well-being and active movement in children by encouraging them to listen to music and dance with a robot. This activity block includes a selection of popular dances, including "Macarena," "Gangnam," and "Tai Chi," each of which is synchronized with a specific song and designed to accommodate a range of body movements. To initiate a dance, children should touch the tactile sensors located on various parts of the robot's body. The interaction flow for this activity block is illustrated in Figure 3-3. By engaging in this activity, children enjoy the benefits of music and movement while developing their perceptual functions (visuospatial, auditory perceptions) (see Table 3-2).

3.4.2 "Songs" Activity Block

The primary aim of the "Songs" activity block is to enhance children's emotional well-being through music, which led to an expansion in the number of songs available from three to twelve. These include "Tanya," "Helper," "Wash Your Hands," "Painter," "Red Apricot," "Spider," "Heroes," "Clock," "Clock," "Spider," "Beautiful," "Mothers,"

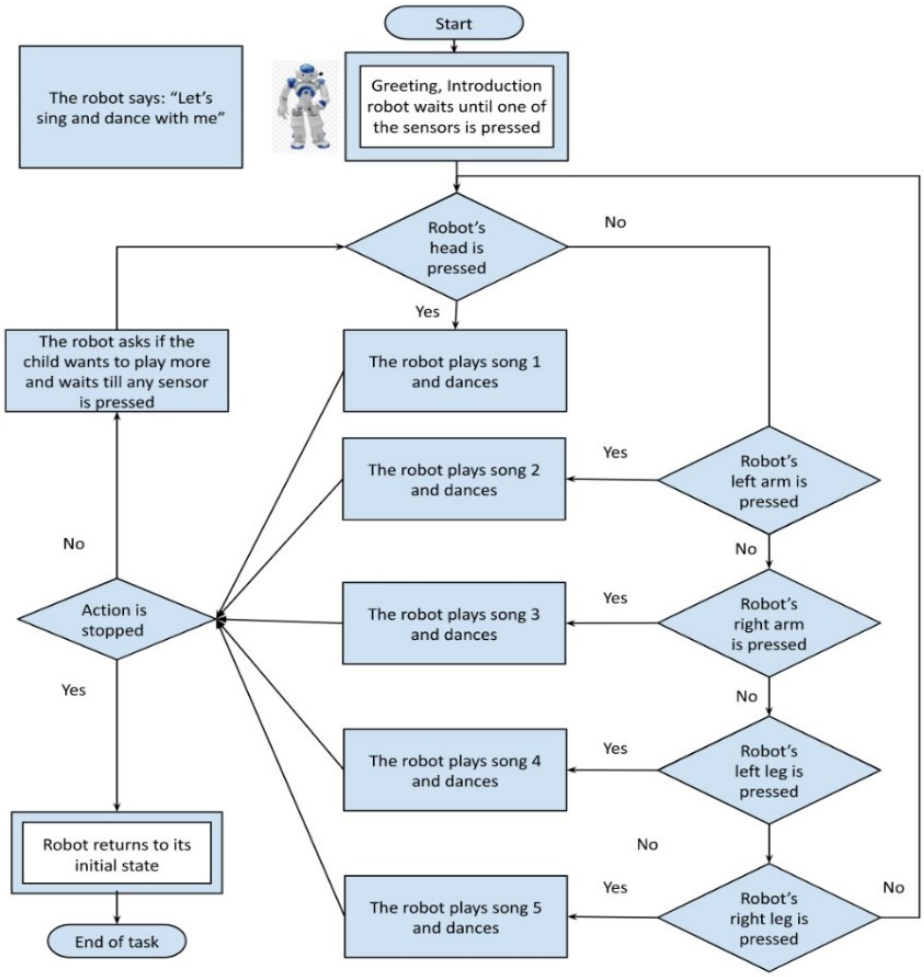

Figure 3-4: Activity diagram for "Songs"

"Fixers," "Maria," and "Mother." In order to move the robot in time with the song's rhythm, a straightforward choreography was created. The music therapist provided these rhythms with different tempos and paces. Some songs were tailored specifically for Kazakh or Russian-speaking children, while others were appropriate for both groups. Tactile sensors located on the robot's head, arms, and feet initiate the playing of these songs, each lasting for approximately one to two minutes. The interaction flow for this activity block is presented in Figure 3-4. By providing a diverse range of songs and choreography, this activity promotes children's cognitive, sensory and motor development in a fun and interactive way (see Table 3-2).

3.4.3 "Action Song" Activity Block

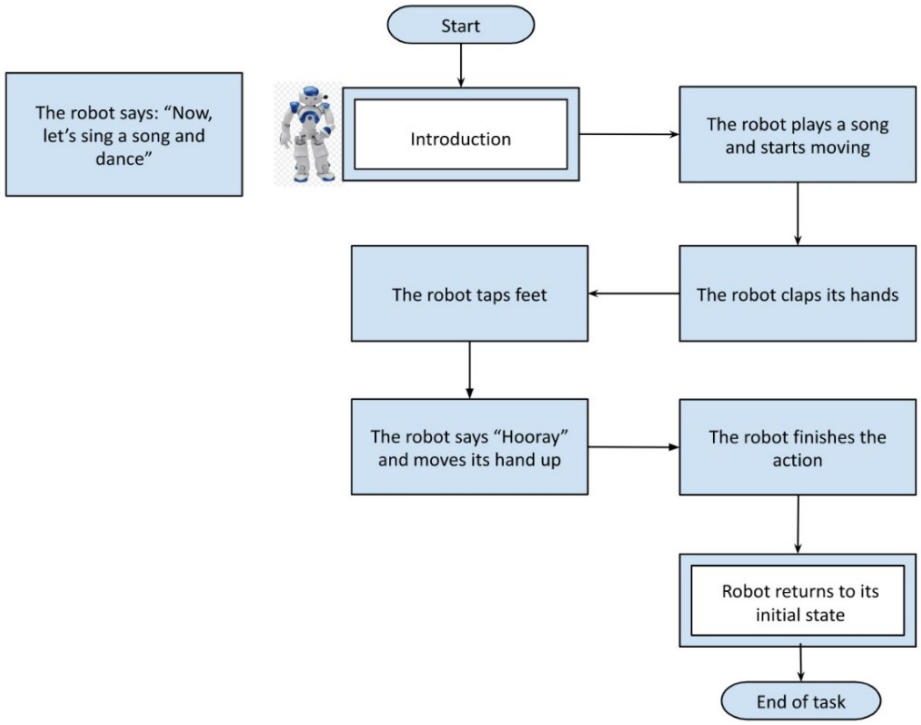

Figure 3-5: Activity diagram for "Action Song"

The "Action Song" activity incorporates the popular song "Clap Your Hands," which encourages children to clap, tap their feet, and shout "hooray" along with the music. The robot demonstrates these movements in sync with the music and encourages children to join in and participate. The activity lasts for approximately 1.5

minutes, providing a brief but engaging experience for children. The interaction flow for this activity block is presented in Figure 3-5. This activity helps to enhance children's verbal and nonverbal skills, perceptual functions, some cognitive development and psychomotor functions in a fun and interactive way (see Table 3-2).

3.4.4 "Storytelling" Activity Block

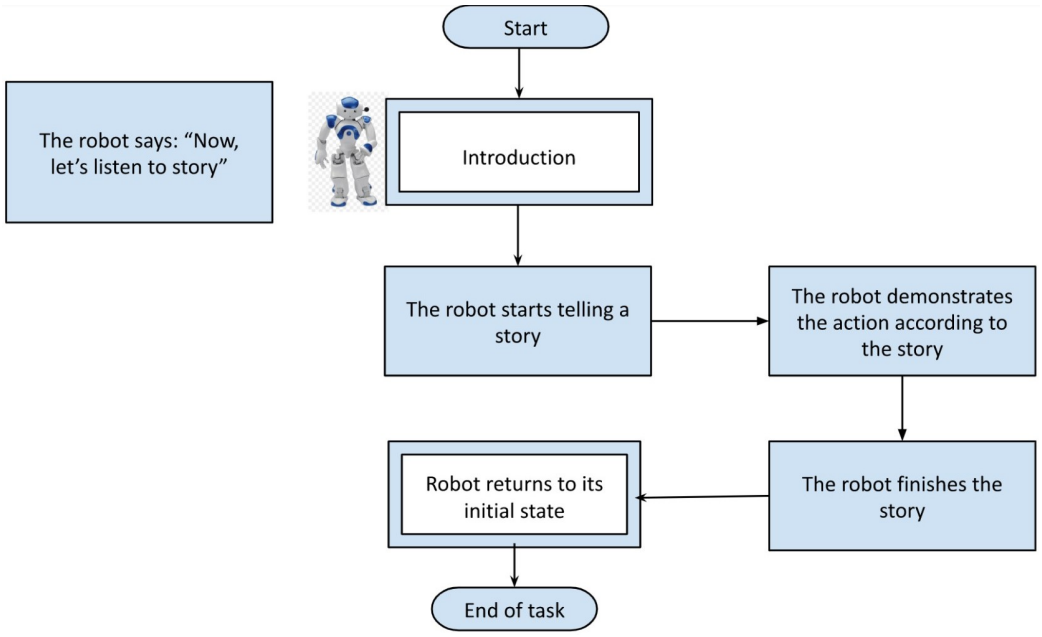

Figure 3-6: Activity diagram for "Storytelling"

In order to enhance children's concentration and imagination skills, we implemented a "Storytelling" activity block. "The Cockerel" fairy tale was added to this block, which already includes "The Bun" and "The Turnip." As the story is told, the robot moves its body parts in sync with the narrative, providing a more engaging and interactive experience for children. The storytelling incorporates animated movements and sounds to further enhance the experience. Moreover, it is supported in two languages. Each story lasts for approximately 1.30-2 minutes, providing a brief but captivating experience for children. The interaction flow for this activity block is presented in Figure 3-6. By incorporating storytelling and movement, this activity promotes children's cognitive and imaginative skills in an entertaining and stimulating way (see Table 3-2).

3.4.5 "Follow Me" Activity Block

The "Follow Me" activity begins with the robot explaining that it has recently been developed and needs to learn how to walk. The robot then asks a child to assist by holding its hand and helping it to walk in the right direction. The child's goal is to guide the robot and help it practice its new skill, while also preventing it from falling over. To stop the robot, the child simply releases its hand. [184]. The interaction flow for this activity block is presented in Figure 3-7. This activity promotes children's social and motor skills by encouraging them to work collaboratively with the robot and practice important physical coordination skills in a fun and engaging way.

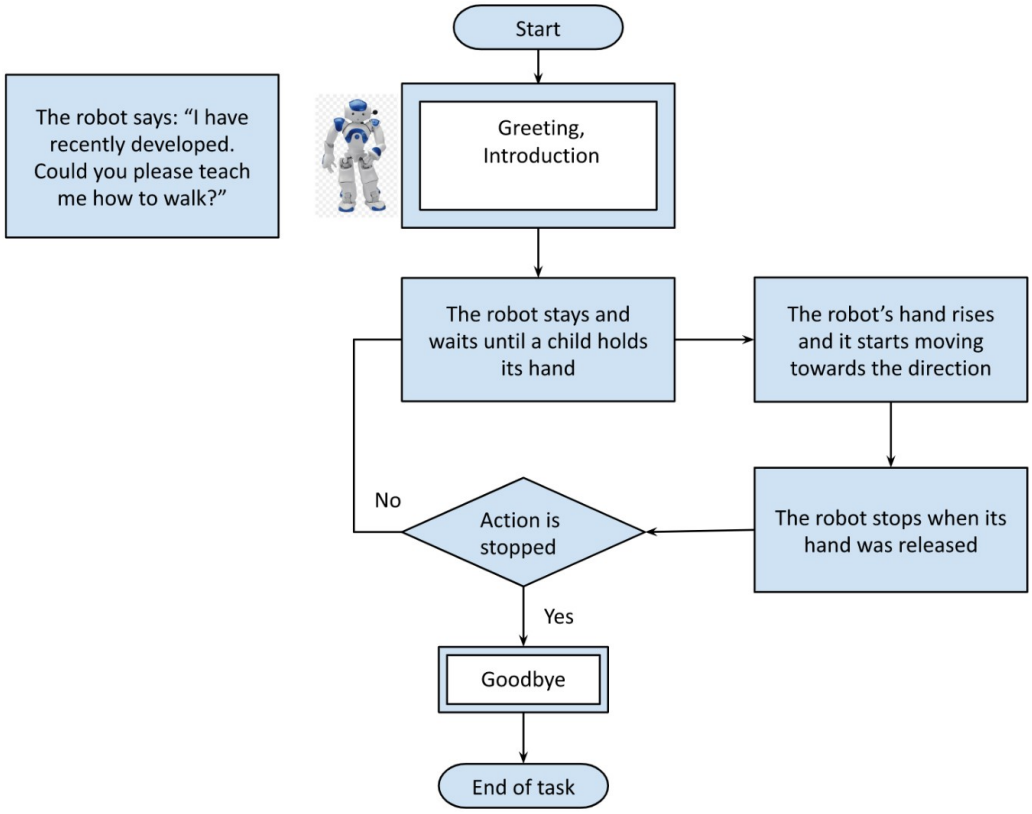

Figure 3-7: Activity diagram for "Follow Me"

3.4.6 "Touch Me" Activity Block

Children with autism may learn various body parts by touching sensors on the robot in the "Touch Me" activity. The robot uses basic instructions like "stroke the blue spot

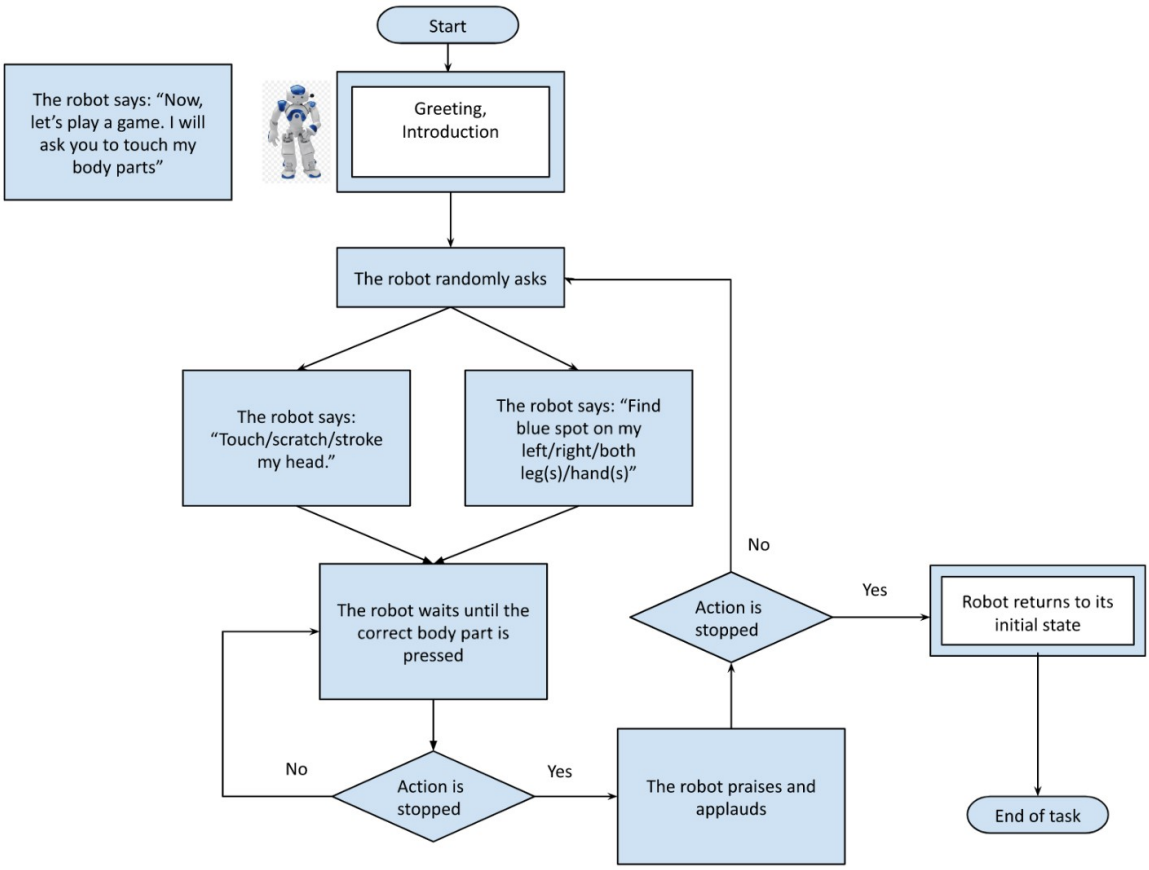

Figure 3-8: Activity diagram for "Touch Me"

on my right hand," "tap my blue toes on my left foot," and "pat on my head" to direct the children to touch various body parts one at a time. The robot replies with praise and applause when the child properly recognizes and touches the requested body. The robot stays stationary if the child touches the wrong section of its body. The interaction flow for this activity is presented in Figure 3-8. By incorporating touch and verbal cues, this activity promotes children's sensory and language development in a playful and interactive way (see Table 3-2).

3.4.7 "Imitations" Activity Block

The robot behaviour was implemented to demonstrate nonverbal and verbal behaviours to enhance imitation skills.

Previously, two activities were created: the "Animals", which included four ani-

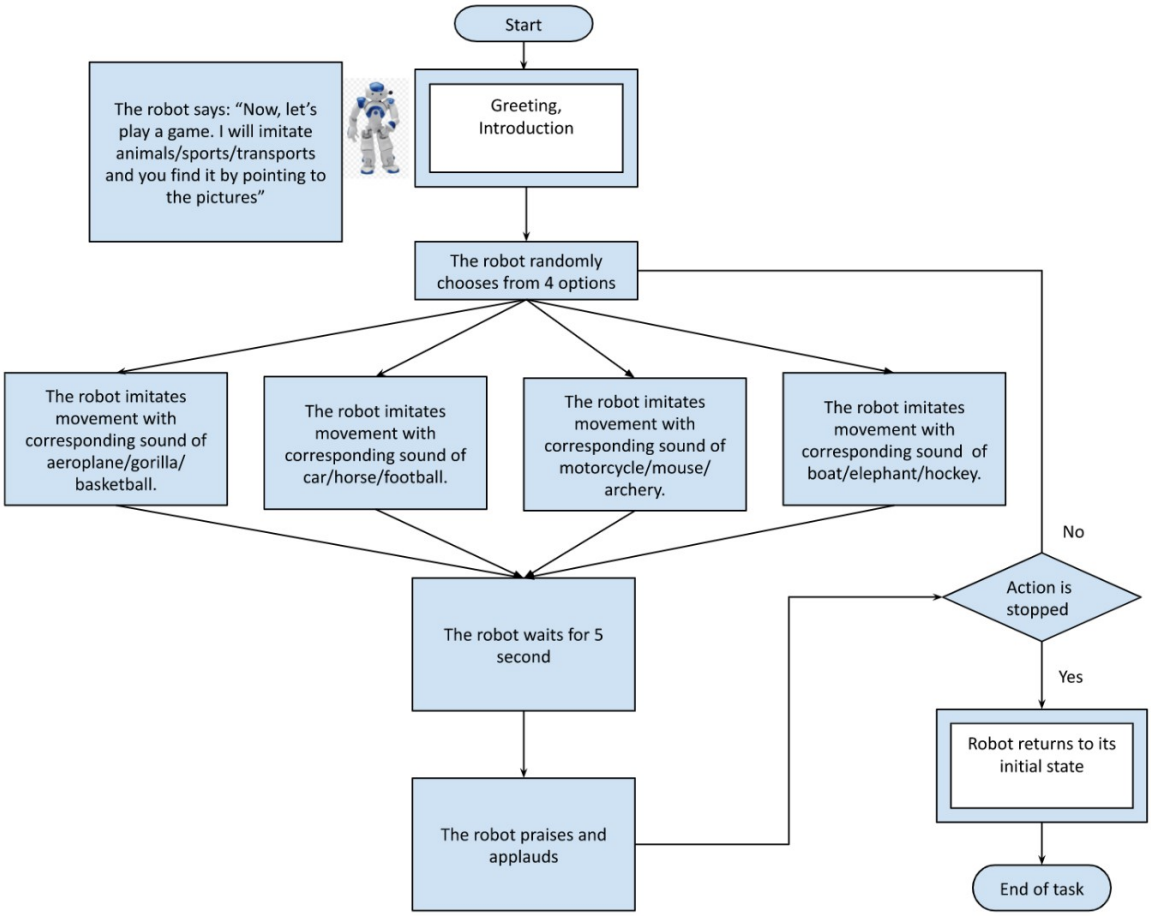

Figure 3-9: Activity diagram for "Imitations"

mated movements of animals such as gorilla, mouse, horse, and elephant; the "Transports", which comprised four animations of transports (boat, car, aeroplane, and motorcycle). To further promote imitation skills, a new "Sports" activity was added, consisting of popular sports like basketball, football, archery, and hockey. Through gestures, movements, and sounds, the robot imitated each sport and encouraged children to replicate the actions. To maintain joint attention, a printed image was presented with each animation. The duration of this activity was halted via the script. The interaction flow for this activity is shown in Figure 3-9. With this type of behaviour, the robot can provide children with engaging opportunities to improve imitation skills in a fun and interactive way (see Table 3-2).

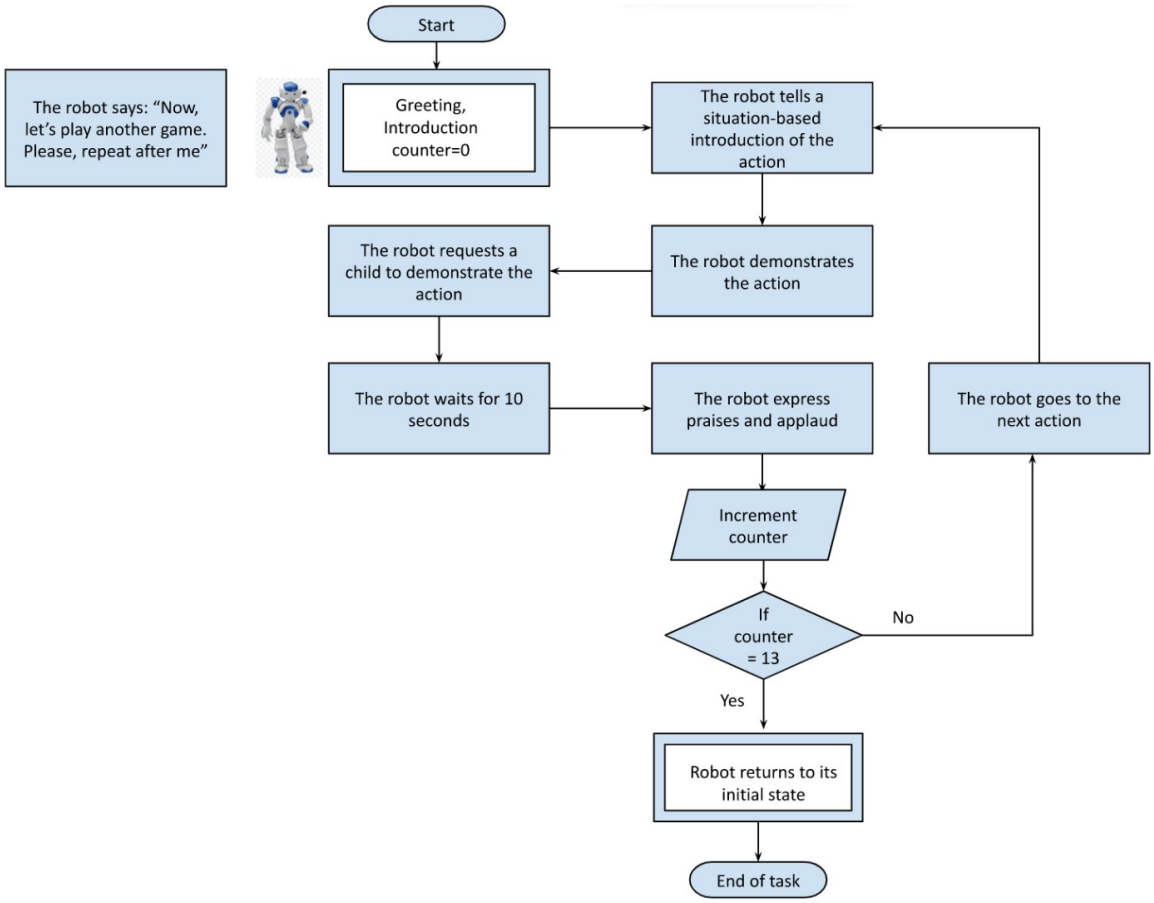

Figure 3-10: Activity diagram for "Social Acts"

3.4.8 "Social Acts" Activity Block

The "Social Acts" activity is designed to help children practice social skills, including turn-taking, imitation, and nonverbal communication. The activity consists of a collection of easy-to-understand nonverbal cues, such as handshakes, high-fives, yawning, peace signs, and other gestures, such as clapping, kissing, etc. To engage children, the robot introduces each action with a situation-based explanation, such as "I yawn when I feel sleepy" or "I clap when I feel happy." The robot asks the child to repeat each move with their parents and therapist after showing it. And waits for 10 seconds for the child to respond. The activity lasts for 4 minutes and includes 13 different nonverbal communicative actions. The interaction flow for this activity is presented in Figure 3-10. This activity promotes children's almost all social and motor skills by encouraging them to repeat and imitate the robot behaviours in a fun and engaging

way (see Table 3-2).

3.4.9 "Emotions" Activity Block

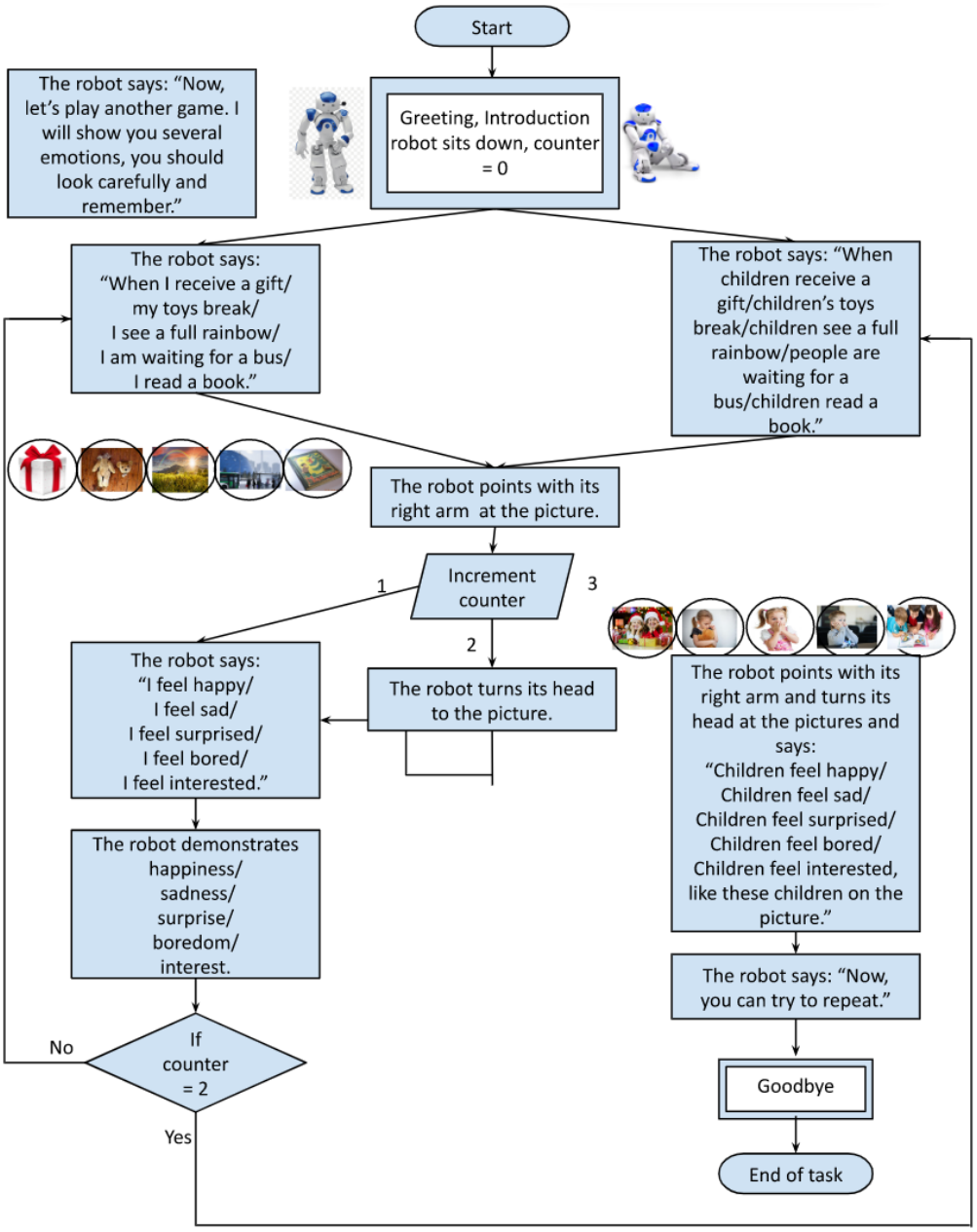

Figure 3-11: Activity diagram for "Emotions"

The "Emotions" activity aims to better promote children's emotional health by helping them recognize affective states and develop joint attention skills. The activity consists of the robot performing five emotions - interested, happy, bored, sad, and

surprised - with the corresponding sound for each of them. Additionally, the activity includes printed photos for each emotion, with two pictures portraying a child showing a particular emotion and the situation that caused it. For example, the robot might point to a picture and say, "This child is bored when he waits for a bus for a long time," and then perform the bored emotion with action by placing its hand on the head. The interaction flow for this activity is presented in Figure 3-11. This activity promotes children's several types of development, such as sensory, cognitive, social and emotional, motor and interaction by encouraging them to work collaboratively with the robot in a fun and engaging way (see Table 3-2).

3.5 Classifications of Activities

Three classifications of play to categorise our activities were adopted: by Piaget (1945), Garvey (1990) and the US National Institute of Play [30] (Figure 3-12). The activities were then grouped into nine activity blocks: "Action Song," "Dances," "Emotions," "Follow Me," "Imitations," "Social Acts," "Songs," "Storytelling," and "Touch Me."

Piaget's (1945) classification of play consists of three categories:

1. play with rules,
2. practice play, which involves visual, tactile, and listening skills,
3. symbolic play, which includes imaginative, make believe activities, and the utilization of absent objects.

Garvey (1990) suggested classifying play behaviours into the following four types:

1. play with objects,
2. play with language, which includes rhymes, sounds, words, noises, etc.,
3. play with motion and interaction, which includes movements like jumping, running, laughing, etc.,

Groups	Activities	Type of play			Social mediation level
		by Piaget	by Garvey	by US National Institute of Play	
Dances	Macarena	●	■	▲▲ ▲	Low
	Ganghnam	●	■	▲▲ ▲	Low
	Thai Chi	●	■	▲▲ ▲	Low
Songs	Mothers	●	■ ■	▲▲ ▲	Medium
	Spider	●	■ ■	▲▲ ▲	Medium
	Clock	●	■ ■	▲▲ ▲	Medium
	Painter	●	■ ■	▲▲ ▲	Medium
	Helper	●	■ ■	▲▲ ▲	Medium
	Fixers	●	■ ■	▲▲ ▲	Medium
	Magdalena	●	■ ■	▲▲ ▲	Medium
	Red Apricot	●	■ ■	▲▲ ▲	Medium
	Heroes	●	■ ■	▲▲ ▲	Medium
	Wash Your Hands	●	■ ■	▲▲ ▲	Medium
	Tanya	●	■ ■	▲▲ ▲	Medium
	Beautiful	●	■ ■	▲▲ ▲	Medium
Action Song	Clap Your Hands	● ●	■ ■ ■	▲▲ ▲ ▲	Medium
Storytelling	The Turnip	● ●	■ ■ ■	▲▲ ▲ ▲ ▲ ▲	Medium
	The Bun	● ●	■ ■ ■	▲▲ ▲ ▲ ▲ ▲	Medium
	The Cockerel	● ●	■ ■ ■	▲▲ ▲ ▲ ▲ ▲	Medium
Follow Me	Follow me	● ●	■ ■ ■	▲▲ ▲ ▲ ▲	High
Touch Me	Touch Me	● ● ●	■ ■	▲▲ ▲ ▲ ▲	High
Imitations	Transports	● ●	■ ■ ■ ■	▲▲ ▲ ▲ ▲ ▲	High
	Animals	● ●	■ ■ ■ ■	▲▲ ▲ ▲ ▲ ▲	High
	Sports	● ●	■ ■ ■ ■	▲▲ ▲ ▲ ▲ ▲	High
Social Acts	Social acts	● ●	■ ■ ■ ■	▲▲ ▲ ▲ ▲ ▲	High
Emotions	Emotions	● ● ●	■ ■ ■ ■	▲▲ ▲ ▲ ▲ ▲	High

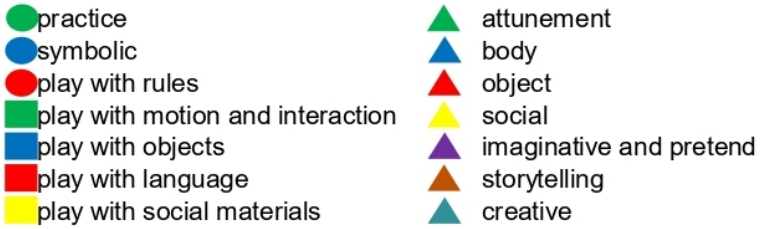

Figure 3-12: Activities categorization by type of play

4. play with social materials, which involves skills such as pretending and make believe.

The US National Institute of Play identifies seven patterns of play:

1. comprised of body play (early rhythmic speech and physical movements),
2. social play (engaging in activities with others),
3. attunement play (joint attention),
4. object play,
5. creative play (the process of generating new ideas and fantasy play),
6. pretend play and imaginative (make-believe activities),
7. storytelling (listening to and narrating stories).

As depicted in Figure 3-12, all activities designed in our study incorporated continuous attunement, social practice, and play with motion, regardless of their level of social mediation. We created activities with diverse levels of social mediation, corresponding to varying levels of behaviour expression. The complexity of the social skills that needed to be practised was indicated by the social mediation levels, which ranged from straightforward repetition to expressing in a complex way of creativity and autonomy. For example, the "Dances" activity had the lowest social interaction complexity, while the "Storytelling" and "Songs" activities represented medium-level social practices. To address social and physical skills and improve emotional skills and we developed dance and music-based activity behaviours. Higher-level cognitive skills were developed in activities like "Emotions," "Social Acts," and "Touch Me" (where a child needs to locate and touch body parts). These activities all included complex goal-oriented behavioural tasks.

These activities promoted emotional and social abilities such as intentional object engagement, empathy, and emotional perception. Children were expected to follow the rules when participating in the "Touch Me" game, which focused on decision-making (which robot sensor to press) and analyzing choices made (when a correct

answer was given, the robot congratulated and applauded; otherwise, it remained silent). All robot behaviours included tactile sensory stimulation, which was specifically designed to improve certain cognitive processes linked to identifying and interpreting sensory stimuli, such as auditory, visuospatial, visual, and tactile functions. A child interacted with things and used tangible materials, including printed images, during the "Emotions" and "Imitations" activity blocks. Moreover, the high-level mediation blocks encouraged creativity, allowing children to try out new behaviours and pick up knowledge from watching the robot.

3.6 Concluding Remarks

This chapter discusses multi-purposeful robot activities that aim to help children with ASD overcome their social, emotional, and motor challenges. These robot activities are based on the ABA principles, which is an evidence-based approach to autism therapy.

Based on the degree to which the robot mediates social interactions between the child and others, these activities are grouped according to their social mediation levels. Therapists and caregivers may select the robot activities that are most appropriate for each child's unique needs and skills by being aware of the various robot activity groups and their levels of social mediation. This method enables a personalized therapeutic strategy that may produce better results for children diagnosed with ASD.

Chapter 4

A Long-Term RAAT

In the study described in this chapter, 11 children with ASD and co-occurring Attention Deficit Hyperactivity Disorder (ADHD) were involved. Each child engaged in a number of therapy sessions with the NAO robot that included a range of activities designed to address each child's unique requirements. To gain a thorough understanding of the children's responses, we employed both qualitative and quantitative analyses, using observations, questionnaires, and video recordings of the interactions. We examined a number of behavioural measures using video recordings, including eye gaze and engagement duration, valence and engagement scores. These measures were extracted from the recordings to provide data on the behaviour of the children during the interventions.

4.1 Ethical Approvals

The Ethics committees of NU and the Republican Children's Rehabilitation Center (RCRC) in Kazakhstan provided their approval to conduct this research. This study followed a previously proposed methodology [184].

The experiments were carried out at the RCRC, which is a leading facility in Kazakhstan that provides rehabilitation services to children with various disabilities, including ASD and ADHD. The center has 365 beds and employs specialists from various fields, including medicine, psychology, pedagogy, social adaptation, art, music

therapy, and physiotherapy. During the 21-day therapy program, children with their parents or caregivers were admitted to the RCRC to receive traditional autism therapy methods.

On the first day of the program, doctors conducted diagnostic assessments of the children to determine their needs and requirements.

The setting of the RCRC provided a rare opportunity to conduct a long-term study with a substantial number of participants. Although parents were not required to attend the therapy sessions, they were encouraged to participate and observe their children. This approach ensured that parents observed the therapy process, which allowed us to conduct interviews with them on their child's progress.

4.2 Recruitment

On the same day, the participating children arrived at the Rehabilitation Center with their parents. The first few days were allotted for them to adjust to the hospital's environment. After this period, a meeting was organized for all parents and their children at the hospital venue. The aim of this meeting was to explain the therapy procedure and introduce the NAO robot, which lasted approximately two hours. It was divided into three parts:

1. In the first part, an overview of the study was provided, which included the purpose, data collection process, potential benefits and risks, intervention procedures, and other related information.

2. In the second part, informed consent forms were sent to the parents.

3. Finally, questions from parents regarding the study were answered.

Parents were given time to read the consent forms on their own. Some parents signed the forms immediately after the meeting, while others returned them the next day.

Furthermore, the contact numbers of the parents were collected to create a group chat. The purpose of this chat was to keep parents informed about scheduling and

for further correspondence. All parents agreed to be included in the group chat. It is important to note that all parents were familiar with each other before the study and had a good relationship.

4.3 Participants

Eleven children (1 girl, 10 boys) between ages 4 and 11 years old, which were diagnosed with ASD participated in the long-term study. Three out of eleven children were verbal. Seven out of the eleven children had a co-occurring diagnosis of ADHD. When the study was conducted, the children's average age was 6.1 years with SD = 2.7. Every child interacted with the NAO robot at least seven sessions on separate days and each session on average lasted for 15 minutes [159].

Table 4.1: The number of sessions attended, demographic information and personal characteristics of each child, such as whether they were verbal or nonverbal, ASD form, ADOS-2 scores, and co-existing ADHD.

ID	**Sessions**	**Gender**	**Age**	**Verbal**	**ASD form**	**ADOS-2**	**ADHD**
C1	10	M	5	-	Moderate	6	-
C2	9	M	10	✓	Severe	9	✓
C3	9	M	7	-	Severe	9	✓
C4	8	M	5	-	Severe	8	✓
C5	8	M	5	-	Severe	8	✓
C6	8	M	5	-	Moderate	6	✓
C7	8	M	5	-	Moderate	7	✓
C8	7	M	10	✓	Severe	9	-
C9	7	M	6	-	Severe	8	-
C10	7	F	5	-	Moderate	5	✓
C11	7	M	4	✓	Moderate	5	-

When the children came to the RCRC, a therapist with approximately seven years of experience conducted an ADOS-2 scoring test to evaluate their communication, play skills and social interaction. The ADOS-2 test [130] is designed to analyze the behaviour of the child during different play scenarios, tasks, and interactions in a sequential manner, with comparative points assessing the severity of autism-related symptoms. A score of 3-4 represents a mild form, 5-7 a moderate form, and 8-10 a severe form [130]. Among the eleven participants, five had a moderate form of autism,

while six had a severe form. The average ADOS-2 test score for all participants was 7.3 (SD = 1.6), indicating moderate to severe autism. Additionally, eight children were nonverbal and only able to speak a few words. To keep the participants' privacy, child IDs were assigned. Table 4.1 displays the participant characteristics as provided by therapists and parents, along with the number of sessions attended.

4.4 Setup

As can be seen in Figure 4-1, therapy sessions were held in a little, empty room with sports mats covering the walls and floor. In order to promote eye contact and free mobility, the NAO robot and the child were both placed on the ground.

The components of the hardware were kept minimal: two NAO robots, a local Wi-Fi router, two cameras, and two PCs, one for a camera to record the interaction and another to control the robots. One recording camera was put close to the child on the mat, while the other was set on the wall to record the whole room. A Wi-Fi router provided wireless connectivity for the robot. The researcher controlled the session by running programs while seated behind the mats at a computer.

The setup was designed to be straightforward and unintrusive, prioritizing the creation of a comfortable and natural environment for the child to engage with the robot.

4.5 Procedure

Each child was offered to engage in up to 10 sessions of RAAT. The planned duration for each session was 15-20 minutes; however, sessions could be stopped if a child lost interest or wished to leave the room. Each child's attendance at sessions varied depending on personal factors, such as a child falling asleep or engaging in other activities, leading to missed sessions.

The NAO robot introduced itself at the start of each session by saying, "Hello, my name is Nao. Today, we will have fun together, dance, and play. Let's start!?"

Figure 4-1: Experimental setup of RAAT

Since all the robot activities were new to the children, the therapist asked them and their parents (in cases where a child was nonverbal) which one they would like to start with. Considering their remarks and the overall characteristics of each child (such as their receptivity to noises), the therapist chose one of the simpler activities like "Touch Me," "Songs," or "Storytelling" for the first session.

During the RAAT sessions, the human therapist selected activities for the robot based on the child's responses to previous activities.

The therapist then instructed the researcher through a command-line interface on which activity to start. According to the child's success on each particular task and interaction with the robot during the session, activities were gradually added. For example, if the robot performed activities like "Dances," "Storytelling," "Songs," and "Touch Me" in the first session, a second song and storytelling activity would be added in the second session. To better understand the unique requirements and preferences of every child, new activities were added to each session. After the third session, the type of activity blocks and their chronological sequence were changed in accordance with the child's level of participation and observed preferences. Some children disliked certain activities, while others preferred specific ones. With every

child, it was impossible to keep the same interaction flow due to the variety of ASD and ADHD conditions. As a result, the activities' sequence was based on the therapist's and the parent's observations.

Following the final session, we conducted interviews with the parents of our participants to comprehend the impact of RAAT on their children. We posed questions that centred on the most pertinent aspects of the therapy. We refrained from asking questions related to the child's personal preferences.

4.6 Type of Sessions

As mentioned earlier, the therapist customized each session based on the children's behaviour and response to the robot activities in order to enhance their engagement and, consequently, maximize the therapeutic benefits. However, some sessions introduced new activities. To strike a balance between familiarity and novelty, all sessions were categorized into two conditions: familiar and unfamiliar sessions. A session was classified as familiar if it primarily consisted of previously experienced activities. On the other hand, if a session largely contained new activities, it was considered an unfamiliar one. As a result, the labelling of all initial sessions was unfamiliar, the labelling changed for the following sessions for each child based on the proportion of familiar to unfamiliar activities.

To establish familiarity with an activity, the therapist monitored the child's past interactions with the robot and evaluated if the child had experienced the particular activity before. Additionally, the therapist observed the child's responses to familiar and unfamiliar activities in order to personalize future sessions.

4.7 Video Coding and Measures

The videos were recorded using a webcam that had a built-in microphone. To ensure the reliability of the coding, two independent researchers coded 50% of the video recordings using the ELAN software (Figure 4-2). To determine the inter-coder agree-

ment, 20% of the data was cross-coded, resulting in an agreement score of 82.6%.

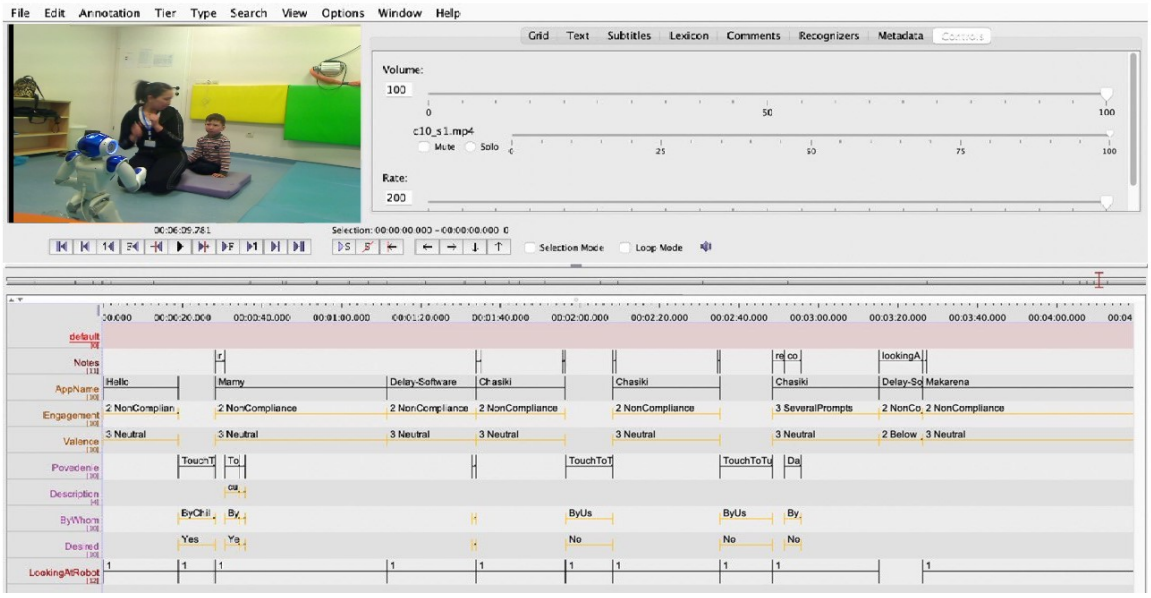

Figure 4-2: The screenshot of labelling videos using ELAN software

To assign engagement and valence scores, we adapted a coding strategy similar to previous works by Kim et al. [107] and Rudovic et al. [179] with combining 0 and 1 scores to 1. As a result, the engagement was rated on a 1-5 Likert scale, with 1 representing the child being evasive and 5 for engagement (see Section 5.4). Similar to the Rudovic et al. [179] we used task-driven coding of engagement, i.e. starting with the target task, e.g., the therapist asking the child to repeat movements and sounds after the NAO robot ("Imitations" activity block), until the child shows different behaviours corresponding to another scoring of the engagement level. Each activity has its own goal of demonstrating how the child can show their engaging behaviours. For example, listening to and looking at the robot while "Dances", "Songs" or "Storytelling" activity blocks, repeating movements and sounds after the robot during "Imitations" and "Emotions" activity blocks, touching the required robot's body parts when it was asked during "Touch Me" activity, and singing and dancing with the robot while "Dances" and "Songs" activity blocks. The detailed annotation of each activity is presented in Table 4.2.

Table 4.2: Coding of engagement measurement for each activity block

Activity block	**Label 1 - Non-compliance**	**Label 2 - Indifference**	**Label 3 - Low engagement**	**Label 4 - Mid engagement**	**Label 5 - High engagement**
Dances	The child is unwilling to dance, not paying attention to the robot, and not reacting to the therapist.	The child repeats dance movements only when the therapist asks more than three times.	The child repeats dance movements only when the therapist asks 2-3 times.	The child repeats dance movements when the therapist or parent dances with the robot.	The child dances with the robot.
Songs	The child is unwilling to listen to a song, not paying attention to the robot, and not reacting to the therapist.	The child repeats dance movements or songs only when the therapist asks more than three times.	The child repeats dance movements or songs only when the therapist asks 2-3 times.	The child repeats a song and dance movements when the therapist or parents repeat it after the robot.	The child sings and dances with the robot.

Table 4.2: Coding of engagement measurement for each activity block (cont.)

Activity block	**Label 1 - Non-compliance**	**Label 2 - Indifference**	**Label 3 - Low engagement**	**Label 4 - Mid engagement**	**Label 5 - High engagement**
Action Song	The child is unwilling to listen to a song, not paying attention to the robot, and not reacting to the therapist.	The child repeats actions from the song only when the therapist asks more than three times.	The child repeats actions from the song only when the therapist asks 2-3 times.	The child repeats an action from the song when the therapist or parents demonstrate it with the robot.	The child demonstrates actions from the song with the robot.
Story-telling	The child is unwilling to listen to a tale, not paying attention to the robot, and not reacting to the therapist.	The child listens to a tale only when the therapist asks more than three times.	The child listens to a tale only when the therapist asks 2-3 times.	The child listens to a tale when the therapist or parents point to the robot to get their attention.	The child listens to a tale and/or repeats the robot's movements.

Table 4.2: Coding of engagement measurement for each activity block (cont.)

Activity block	**Label 1 - Non-compliance**	**Label 2 - Indifference**	**Label 3 - Low engagement**	**Label 4 - Mid engagement**	**Label 5 - High engagement**
Follow Me	The child is unwilling to hold the robot's hand and walk, not paying attention to the robot, and not reacting to the therapist.	The child holds the robot's hand only when the therapist asks more than three times.	The child holds the robot's hand and walks only when the therapist asks 2-3 times.	The child holds the robot's hand and walks with the help of the therapist or parents.	The child holds the robot's hand and walks.
Touch Me	The child is unwilling to touch the robot's body part, not paying attention to the robot, and not reacting to the therapist.	The child touches the requested robot's body part only when the therapist asks more than three times.	The child touches the requested robot's body part only when the therapist asks 2-3 times.	The child touches the robot's body part with the help of a parent or after being pointed to by the therapist.	The child touches the requested robot's body part.

Table 4.2: Coding of engagement measurement for each activity block (cont.)

Activity block	**Label 1 - Non-compliance**	**Label 2 - Indifference**	**Label 3 - Low engagement**	**Label 4 - Mid engagement**	**Label 5 - High engagement**
Imitations	The child is unwilling to listen to the robot, not paying attention to the robot, and not reacting to the therapist.	The child repeats movements and/or sounds only when the therapist asks more than three times.	The child repeats movements and/or sounds only when the therapist asks 2-3 times.	The child repeats movements and/or sounds with the help of the therapist or parent, like demonstrating with the robot.	The child repeats movements and/or sounds after the robot's demonstrations.
Social Acts	The child is unwilling to watch the robot, not paying attention to the robot, and not reacting to the therapist.	The child repeats movements only when the therapist asks more than three times.	The child repeats movements only when the therapist asks 2-3 times.	The child repeats movements when the therapist or parents demonstrate them with the robot.	The child repeats movements after the robot's demonstrations.

Table 4.2: Coding of engagement measurement for each activity block (cont.)

Activity block	**Label 1 - Non-compliance**	**Label 2 - Indifference**	**Label 3 - Low engagement**	**Label 4 - Mid engagement**	**Label 5 - High engagement**
Emotions	The child is unwilling to watch the robot's emotions, not paying attention to the robot, and not reacting to the therapist.	The child repeats emotions only when the therapist asks more than three times.	The child repeats emotions only when the therapist asks 2-3 times.	The child repeats emotions when the therapist or parents demonstrate them with the robot.	The child repeats emotions after the robot's demonstrations.

Table 4.2: Coding of engagement measurement for each activity block (cont.)

After the engagement episodes were coded in the videos, each episode was further coded in terms of valence. ~~Likewise,~~Valence was rated on a 1-5 Likert scale, with 1 denoting the child expressing negative emotions and 5 signifying positive emotions: 1 - clear signs of experiencing negative emotions like crying, being unhappy, angry, visibly upset, showing dissatisfaction, frightened; 2 - displaying feelings of sadness or boredom; 3 - remaining neutral; 4 - showing interest; and 5 - exhibiting happiness or excitement. For example, the episodes were coded with high negative valence (1) in cases when the child was crying, screaming, or showing fear or anxiety during the activities. Also, the valence score was assigned to 2, when the child was bored and disengaged during the interaction with the NAO robot. When the child did not show any negative or positive emotions, that episode was coded as 3. Furthermore, the valence score was assigned to 4, when the child showed interest, like curiosity, smiling, and paying close attention to the robot. The very positive valence (5) was

coded when the child was happy and engaged in the interaction with the NAO robot.

Overall, we coded four measurements: $S_nEngagement$, $S_nValence$, $S_nEyeGazeTime$, and $S_nEngagementTime$. While $S_nEyeGazeTime$ and $S_nEngagementTime$ were computed as percentages related to the whole length of the session (for example, engagement time of 4 minutes out of a 10-minute session resulting in a value of 10%), engagement and valence scores were coded in relation to the timing of activities. For sessions 1 through n, all measurements contained n variables. As in one session child showed different labels of engagement and valence scores, we calculated the mean of these measurements.

There were two different types of sessions during the intervention: familiar and unfamiliar. We calculated the average of all data for both known and unfamiliar sessions individually, indicating the mean scores for each session type, in order to assess the efficacy of the therapy.

4.8 Results

We performed a series of one-way repeated measures ANOVA (RM ANOVA) with a Greenhouse-Geisser correction (GGC) on data from 11 participating children to look at possible variations in valence and engagement scores, engagement duration, and eye gaze time across sessions. We compared within-subject changes between different sessions using statistical methods.

4.8.1 Comparison Between Sessions

We tested H1 by comparing the measurements of all children across sessions to determine if there were any significant differences during the therapy. The measurements included $S_nEngagement$, $S_nValence$, $S_nEngagementTime$, and $S_nEyeGazeTime$. The findings showed that there was no statistically significant variation between the children's engagement levels between sessions: $F(6, 60) = 3.0, p = 0.05$.

We can observe from Table 4.3, that engagement and valence scores, as well as engagement and eye gaze duration, varied throughout the course of the 10 sessions.

Table 4.3: For each session, the mean values for scores of engagement and valence, as well as the durations of engagement and eye gaze

Measurement	**s1**	**s2**	**s3**	**s4**	**s5**
Engagement score	2.82 ± 0.63	3.24 ± 0.82	2.86 ± 0.65	3.12 ± 0.67	3.34 ± 0.67
Valence score	3.21 ± 0.41	3.36 ± 0.43	3.34 ± 0.53	3.45 ± 0.45	3.35 ± 0.42
Engagement Time	56.57 ± 26.3	59.69 ± 24.41	58.02 ± 22.83	69.25 ± 21.58	75.64 ± 20.69
Eye Gaze Time	53.83 ± 21.17	70.72 ± 18.53	69.08 ± 18.87	70.3 ± 18.98	69.7 ± 10.49
Measurement	**s6**	**s7**	**s8**	**s9**	**s10**
Engagement score	2.93 ± 0.56	2.82 ± 0.66	2.93 ± 0.86	3.1 ± 0.83	3.53 ± 0.41
Valence score	3.31 ± 0.27	3.38 ± 0.42	3.48 ± 0.54	3.34 ± 0.47	3.58 ± 0.26
Engagement Time	66.4 ± 17.78	59.9 ± 20.88	73.13 ± 24.03	75.88 ± 19.54	78.38 ± 20.14
Eye Gaze Time	68.3 ± 18.46	42.56 ± 12.79	71.24 ± 25.56	73.8 ± 19.66	79.4 ± 12.28

However, compared to the first session, there was a modest rise in these values in the last session. This implies that over long-term study, children with autism can be engaged to communicate with robots.

4.8.2 Familiar vs Unfamiliar

In order to assess Hypothesis 1, we conducted RM ANOVA using a GGC on a sample of 11 children. The findings revealed statistically significant differences in the duration of engagement as detailed below: the mean engagement score during unknown sessions was notably lower (2.94 ± 0.22) in comparison to known sessions (3.11 ± 0.24): $F(1.507, 10.546) = 4.678, p = 0.043$; the average engagement length in unfamiliar sessions was significantly reduced (69.98 ± 10.1) when compared to familiar sessions (70.87 ± 8.51): $F(2.384, 16.689) = 3.446, p = 0.049$; the mean eye gaze duration for unfamiliar sessions was considerably lower (65.57 ± 8.84) relative to familiar sessions (75.2 ± 6.82): $F(2.087, 14.609) = 5.232, p = 0.018$. Figure 4-3 visually displays these results for familiar and unfamiliar sessions.

4.8.3 Observations

Over the last few sessions, there has been a noticeable shift in the activities that each child enjoys the most. For instance, C11 showed a preference for "Touch Me"

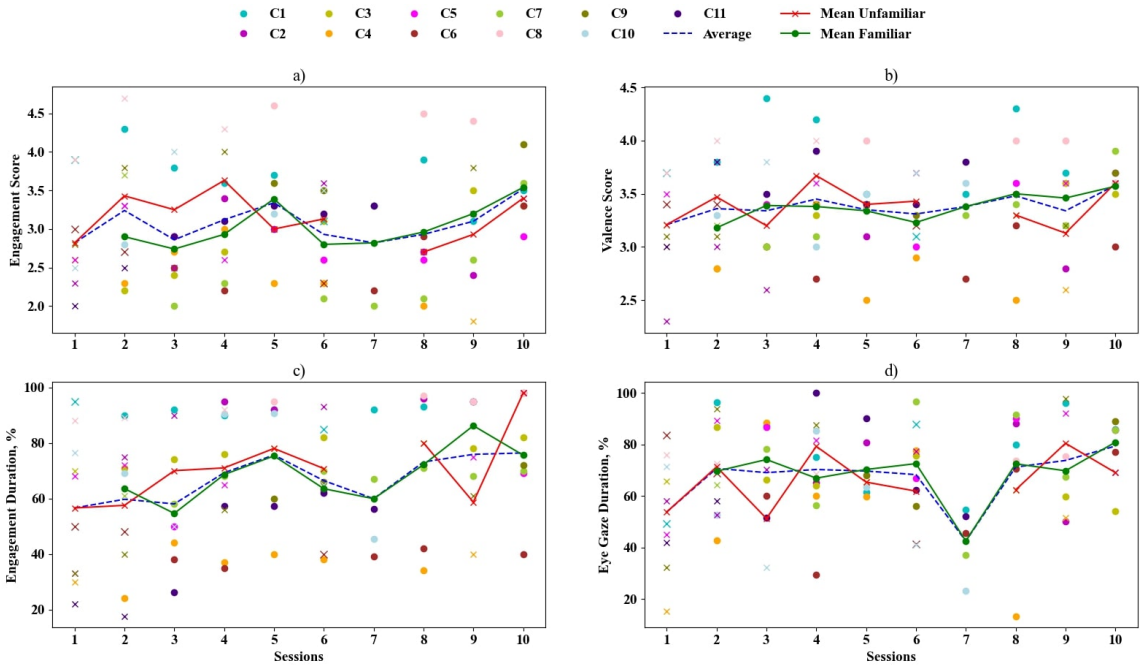

Figure 4-3: For each child in every session, scores of engagement and valence, as well as the duration of engagement and eye gaze. Circles denote familiar sessions, while crosses (x) indicate unfamiliar ones. Each child is represented by a distinct colour.

and "Storytelling," while not being as fond of "Songs" and "Dances." Similarly, C2 did not enjoy song and dance activities as much as "Storytelling," "Emotions," and "Imitations." However, C2 did not have a strong preference for "Touch Me" compared to C11.

Below, we provide more detailed observations of each child.

C1: This five-year-old boy attended ten sessions and played various activities with the robot, as listed in Figure 4.4. Initially, C1 was hesitant to touch the robot, but after watching it dance for five minutes, he became more comfortable and started to interact with the tactile sensors to initiate the "Songs" activity. C1 quickly learned which body parts to press for the "Clock" dance, launching it 14 times during his first session and 47 times throughout the experiment. He also enjoyed the "Spider" activity (six times) and the "Mother" activity (three times). C1 actively engaged in the dance movements and imitated the robot with the help of the researcher. It is important to note that C1 is nonverbal but managed to say his first words, "Tick-Tack," during the

third session and "NAO" during the seventh session. Personalizing the therapy for C1 involved selecting imitation activities such as "Dances," "Transports," "Emotions," and "Animals," while the "Storytelling" activity didn't hold his attention for long.

Table 4.4: Activity blocks played for child C1

Session	**Activity blocks and their frequencies**
S1	Songs(x23)
S2	Songs(x3), TouchMe(x1)
S3	TouchMe(x1), Songs(x10), Imitations(x1)
S4	Songs(x12), Imitations(x1)
S5	Imitations(x2), Songs(x4), Storytelling(x1)
S6	Songs(x7), Emotions(x1)
S7	Emotions(x1), Imitations(x1), Songs(x7)
S8	Imitations(x2), Emotions(x1), Songs(x7)
S9	Emotions(x2), Storytelling(x1), Imitations(x1), Songs(x4)
S10	Emotions(x2), Imitations(x1), Songs(x9)

C2: This ten-year-old boy attended nine sessions, and the list of attended sessions and activities played with the robot are presented in Figure 4.5. C2 has been diagnosed with both ASD and ADHD. Initially, C2 was scared of the new environment, and the robot didn't catch his attention. However, during the second session, he responded well to the "Storytelling" activity and repeated the robot's actions during the stories. After that, he greeted the robot and remembered its name. Personalizing the therapy for C2 involved starting the sessions with the "Storytelling" activity blocks. During the nine sessions, this activity was played 16 times. He began engaging in task-based activities with the robot during the third session. He learned the sounds of animals and transports and started matching them with images during the "Imitations" activity block. His favourite animal was a mouse, and his least favourite was a horse. During the fourth session, he said goodbye to the robot, "See you again, bye!" Moreover, he learned five emotions during the "Emotions" activity and pointed to the corresponding pictures when the NAO robot started talking. However, as he was sensitive to sounds, the "Songs" activity was only played five times for C2.

Table 4.5: Activity blocks played for child C2

Session	Activity blocks and their frequencies
S1	Songs(x2)
S2	Storytelling(x1), TouchMe(x1), Imitations(x1)
S3	Storytelling(x3), Imitations(x2)
S4	Storytelling(x2), Imitations(x2), Songs(x1), Emotions(x1)
S5	Storytelling(x2), Imitations(x2), Emotions(x1), Songs(x1)
S6	Storytelling(x3), Imitations(x2), Emotions(x1)
S7	Storytelling(x2), Imitations(x2), TouchMe(x1), Songs(x1), Emotions(x1)
S8	Storytelling(x2), Emotions(x1)
S9	Storytelling(x1), Emotions(x2), Imitations(x2)

C3: This seven-year-old boy attended nine sessions, and the list of attended sessions and activities played with the robot are presented in Figure 4.6. C3 was diagnosed with both ASD and ADHD. On his first day, C3 was slightly afraid of the robot. However, during his second session, he enjoyed watching the robot dance more than interacting with it and following its instructions. He only engaged in task-based activities with the prompt of his mother. However, the "Emotions" activity never interested C3. In the last three sessions, he showed a willingness to dance "Gangnam Style."

Table 4.6: Activity blocks played for child C3

Session	Activity blocks and their frequencies
S1	Storytelling(x2), Songs(x7)
S2	Storytelling(x2), Touch Me(x1), Songs(x4)
S3	Touch Me(x1), Dances(x1), Songs(x3), Imitations(x1)
S4	Touch Me(x1), Storytelling(x1), Dances(x3), Imitations(x1)
S5	Storytelling(x1), Songs(x3), Dances(x3), Imitations(x1)
S6	Storytelling(x1), Dances(x3), Touch Me(x1), Songs(x4)
S7	Touch Me(x1), Imitations(x1), Songs(x2), Storytelling(x1), Dances(x2)
S8	Emotions(x1), Songs(x4), Dances(x1)
S9	Emotions(x1), Songs(x6), Dances(x1)

C4: This five-year-old boy attended eight sessions and is diagnosed with both ASD and ADHD. Figure 4.7 shows the list of activities played with the robot during his sessions. At first, the robot seemed like a toy to C4, and he interacted with it by moving its arms and legs and trying to lift it. However, he quickly became distracted

during task-based activities. During the "Dances" and "Songs" activities, C4 would cover his ears and sometimes scream, but he would calm down when the robot stopped and sat down. On one occasion, C4 showed aggressive behaviour towards the robot by trying to push it over.

Despite these challenges, C4 did show some preferences for certain activities. During his last two sessions, he enjoyed watching the "Heroes" and "Painter" dances. It may be worth exploring other dance activities or finding ways to make the task-based activities more engaging for C4 in future sessions.

Table 4.7: Activity blocks played for child C4

Session	**Activity blocks and their frequencies**
S1	Songs(x1), Touch Me(x2), Imitations(x1), Dances(x1)
S2	Touch Me(x1), Imitations(x1), Dances(x1)
S3	Touch Me(x1), Songs(x2), Storytelling(x1)
S4	Imitations(x1), Touch Me(x1), Songs(x2), Storytelling(x1)
S5	Touch Me(x1), Emotions(x1), Songs(x2), Storytelling(x1)
S6	Imitations(x2), Songs(x3)
S7	Storytelling(x1), Songs(x4), Imitations(x1)
S8	Emotions(x1), Songs(x8), Storytelling(x1)

C5: This five-year-old boy attended eight sessions, and the list of attended sessions and activities played with the robot are presented in Figure 4.8. The child is diagnosed both with ASD and ADHD. During the first session, C5 did not interact with the robot and played task-based games only with his mother's help. However, in the second session, he smiled when the robot danced and showed interest in touching it. With his mother's support, C5 engaged in task-based activities such as "Transports" and "Emotions" and was able to match the pictures of vehicles and facial expressions with sounds. By the fifth session, C5 started to follow the robot's instructions and imitate its movements during the Dance with Me" activity. He also showed improvement in his attention span and was able to complete longer activities such as Storytelling" with minimal support from his mother. Overall, the therapy showed promise in improving C5's social and cognitive skills.

Table 4.8: Activity blocks played for child C5

Session	Activity blocks and their frequencies
S1	Imitations(x1), Songs(x1)
S2	Songs(x4), Touch Me(x1), Dances(x2), Imitations(x1)
S3	Imitations(x1), Storytelling(x1), Songs(x1), Touch Me(x1)
S4	Imitations(x1), Touch Me(x1), Songs(x6)
S5	Touch Me(x1), Imitations(x1), Songs(x5)
S6	Touch Me(x1), Songs(x4), Imitations(x1), Storytelling(x1)
S7	Emotions(x1), Songs(x4), Touch Me(x1)
S8	Songs(x10)

C6: This is a male child, aged five years, who attended eight sessions. The list of attended sessions with played activities with the robot is presented in Figure 4.9. The child is diagnosed with both ASD and ADHD. During the therapy sessions, the child showed more interest in touching the robot rather than interacting with and following the robot's instructions. However, he seemed to lose interest in the robot during the last three sessions and became distracted by the objects in the room. Despite this, he showed curiosity in investigating the robot by moving its head and fingers. C6 also tried showing pictures to the robot by bringing paper close to its eyes. Additionally, C6 also tried to engage the robot by showing pictures to it. He kept close to his mother during the sessions but still enjoyed watching the "Fixers" and "Painter" songs.

Table 4.9: Activity blocks played for child C6

Session	Activity blocks and their frequencies
S1	Songs(x8), Storytelling(x1)
S2	Emotions(x1), Imitations(x1), Dances(x1), Touch Me(x1)
S3	Touch Me(x1), Songs(x2), Storytelling(x1), Imitations(x1)
S4	Touch Me(x1), Storytelling(x1), Songs(x2), Dances(x1)
S5	Songs(x10), Imitations(x1)
S6	Songs(x5), Touch Me(x1)
S7	Songs(x4), Storytelling(x1), Imitations(x1), Touch Me(x1)
S8	Songs(x8)

C7: The participant, in this case, is a five-year-old boy diagnosed with both ASD and ADHD who attended eight therapy sessions. Figure 4.10 presents the list of

attended sessions with the activities played with the robot. During all sessions, C7 was very quiet and did not actively engage in task-based activities with the robot. However, he showed an interest in exploring the robot by touching and moving its arms and fingers and pressing its chest button. The child enjoyed watching the robot dance and listening to storytelling activities. He even danced with the robot during the "Mother" song with the prompt of the parent. Additionally, from the fourth session, C7 showed an interest in playing the "Animals" and "Transports" activities with the robot.

Table 4.10: Activity blocks played for child C7

Session	**Activity blocks and their frequencies**
S1	Storytelling(x2), Touch Me(x1), Imitations(x1), Songs(x3)
S2	Songs(x4), Imitations(x1), Storytelling(x1)
S3	Storytelling(x1), Imitations(x2)
S4	Imitations(x2), Songs(x7), Storytelling(x1)
S5	Imitations(x2)
S6	Imitations(x1), Songs(x8), Touch Me(x2), Storytelling(x1)
S7	Imitations(x2), Emotions(x1), Songs(x7)
S8	Emotions(x1), Songs(x8)

C8: This is a male child, aged ten years, who attended seven sessions. The list of attended sessions with played activities with the robot is presented in Figure 4.11. This child demonstrated an immediate interest in the robot and showed good engagement with the activities. According to the video recording, he spent an average of 12 minutes without being distracted during each session. He listened attentively and responded positively to the robot's emotions. Starting from the third session, the child actively participated in all activities and enjoyed recalling emotions learnt in the previous sessions. He did not show much interest in playing the "Songs" activity block, but enjoyed listening to the "Storytelling" and even repeated words after the robot. He also used pictures while listening to fairy tales, which was unique compared to other children. From the third session, he started to request the robot to play the "Gangnam Style" song from the beginning of each session and was delighted to dance with the robot. At the end of each session, he gave a high five to the robot, indicating

a positive attitude towards the robot.

Table 4.11: Activity blocks played for child C8

Session	Activity blocks and their frequencies
S1	Songs(x8), Storytelling(x1)
S2	Touch Me(x1), Imitations(x1), Storytelling(x1), Dances(x1)
S3	Emotions(x1), Dances(x1), Storytelling(x1)
S4	Storytelling(x2), Touch Me(x1), Dances(x1)
S5	Imitations(x1), Storytelling(x3), Songs(x3), Dances(x1)
S6	TouchMe(x1), Storytelling(x2), Emotions(x1), Dances(x1), Songs(x4)
S7	Wash Your Hands(x2), Emotions(x1), Dances(x2), Storytelling(x1)

C9: This six-year-old boy attended seven sessions, as listed in Figure 4.12. He demonstrated immediate compliance with the robot and repeated actions after it with his mother's help. The child actively played "Touch Me" and correctly recognised all body parts. Additionally, he showed enjoyment when the robot danced and liked touching its head, hands, fingers, and back. He accurately matched all emotions with pictures of situations and participated actively in all activities, with his favourite being the "Mother" song. The child remained patient during delays due to technical issues but would scream when he got bored. Overall, he was quiet and attentive during the sessions.

Table 4.12: Activity blocks played for child C9

Session	Activity blocks and their frequencies
S1	Storytelling(x1), Songs(x3)
S2	Touch Me(x1), Emotions(x1), Songs(x3), Dances(x1)
S3	Imitations(x2), Songs(x5)
S4	Touch Me(x1), Storytelling(x1), Songs(x7), Dances(x1)
S5	Touch Me(x1), Songs(x11)
S6	Emotions(x2), Songs(x9)
S7	Emotions(x1), Songs(x12), Imitations(x1)

C10: This five-year-old girl attended seven sessions. The list of attended sessions with played activities with the robot is presented in Figure 4.13. The child is diagnosed both with ASD and ADHD. She almost does not speak, her extremely rare

speech is monosyllabic, but she understands instructions and performs them. She responds to her name after calling her two or three times. During the therapy sessions, the child showed a great interest in the robot and its activities. She seemed to enjoy listening to the robot's voice and music and was able to follow simple instructions given by the robot, such as raising her hands, clapping, or dancing.

During the first few sessions, the child seemed a bit shy and reserved, but she gradually started to show more engagement with the robot. She liked to touch the robot and explore its different features, such as buttons and sensors. She also enjoyed playing with the ball pool and other toys in the therapy room.

Table 4.13: Activity blocks played for child C10

Session	**Activity blocks and their frequencies**
S1	Dances(x3)
S2	Dances(x6), Touch Me(x3)
S3	Dances(x9), Songs(x5)
S4	Dances(x5), Touch Me(x3)
S5	Dances(x2), Songs(x5)
S6	Songs(x9), Dances(x4), Storytelling(x1)
S7	Songs(x5), Storytelling(x2)

C11: This four-year-old boy child attended seven therapy sessions, and the list of activities he engaged in with the robot is presented in Figure 4.14. The child has difficulties responding to his name immediately and maintaining eye contact for a long time. His speech is limited to simple sentences, although he can describe actions in pictures. During the sessions, C11 struggled with sitting in one place for an extended period and demonstrated undesirable behaviour such as screaming, shouting, or asking questions such as "Why?" "What for?" or "How so?". Despite these challenges, the child showed some progress during the therapy sessions. He was able to engage in various activities with the robot, such as playing the "Animals" and "Transports" games, and he seemed to enjoy dancing with the robot during the "Mother" song. However, he still struggled to maintain attention and focus during the sessions.

Table 4.14: Activity blocks played for child C11

Session	Activity blocks and their frequencies
S1	Dances(x3), Touch Me(x3)
S2	Touch Me(x1)
S3	Touch Me(x4), Storytelling(x2), Songs(x1)
S4	Touch Me(x3), Storytelling(x2)
S5	Touch Me(x4), Storytelling(x4), Songs(x3), Dances(x2)
S6	Touch Me(x2), Storytelling(x1)
S7	Storytelling(x2), Touch Me(x3), Imitations(x4)

4.9 Interviews: Feedback and Recommendations

In total, two therapists and five parents provided their feedback and recommendations to enhance robot behaviours during RAAT. It is important to note that four parents (P1, P4, P5, and P7) were unable to offer suggestions for improvement. Additionally, the audio recordings of P10 and P11 were saved in the wrong format and could not be accessed.

Although the parents expressed satisfaction with the therapy overall, they recommended that the robot exhibit more active behaviour. For example, P2 suggested increasing the robot's interactivity to make it more engaging: "There should be more live communication. Only children interested in the robot interact with it" (P2). Parent, P3, suggested implementing object recognition and more dynamic behaviours for the robot, such as kicking a ball and using children's names.

P8 suggested adding more rhymes and melodies for verbal children who can repeat after the robot. A recommendation from P6 was to create more mentally stimulating and educational activities, like using visual cards to display various colours and shapes. Similarly, P9 suggested that the robot could teach counting or differentiate between colours.

The RAAT was a novel experience for both therapists. T1 therapist provided her opinion in the perspective on the CRI setting, which states that "children should be left alone to observe their behaviours from outside." T2 therapist highlighted the importance of not forcing children to communicate with a robot. As there were parents, who were pushing their children to interact with the robot. The T2 therapist

also pointed out that "children with autism do not like to wait," which could have negative consequences for their actions throughout the intervention. The therapists proposed to implement activities similar to kicking a ball and playing the xylophone. It might enhance the turn-taking and imitation skills of children. The importance of a long-term triadic relationship involving a robot, a child with autism, and a typically developed child was also acknowledged.

4.10 Discussion

In general, RAAT intervention indicates that children with autism sustained their engagement with the robot during a long-term interaction. While we observed only a small difference in engagement results, the intervention was successful in maintaining the children's engagement over multiple sessions. Additionally, we found that valence scores and eye gaze time remained consistent across the labelled sessions, suggesting that the children re-engaged in the interaction.

The H1 hypothesis investigated the relationship between engagement and children's preference for familiar or unfamiliar activities. We found that the mean of engagement score, engagement duration and eye gaze duration during familiar sessions were significantly higher than during unfamiliar sessions, allowing us to accept the hypothesis. In line with prior research [37, 107], children's positive emotions for specific activities can be regarded as indicators of their engagement. In other words, when we employed preferred activities, as determined by the human therapist, the children were more positive, focused, and engaged. Therefore, it seems advantageous to create a personalized autism intervention that takes into account each child's unique needs and abilities, given the considerable individual differences observed among children diagnosed with autism participating in RAAT.

Our study yielded interesting findings regarding long-term engagement in autism therapy involving CRI. Firstly, we did not observe a significant increase in the children's engagement levels despite the use of versatile activities. Secondly, we found that children were more engaged with familiar activities than unfamiliar ones, indi-

cating a preference for activities that they were comfortable with. This finding contradicts previous research on typical CRI, which suggests that incorporating novelty and variation into an interaction can prevent predictability and boredom [45, 101]. However, for children with autism, it appears that a preference for routine and sameness, and aversion to change, could explain why they preferred familiar activities [193]. Thus, this core feature of autism may explain why children with autism were more preferred to familiar activities.

A way to improve engagement and the efficacy of autism therapy using robots is by developing autonomous robotic systems. Prior researchers [91, 137, 145] have focused on creating adaptive and autonomous robots that can enhance RAAT sessions by identifying the unique and social cues of the children and offering them individualized and thorough experiences. Due to ethical and practical constraints, however, there are concerns regarding the robot's long-term ability to respond to the child in an acceptable manner [235]. Despite these restrictions, it is essential to keep looking into new play-scenarios and applications that can positively influence autism research. To further support the role of the robot as an assistant or mediator in autism therapy, it is necessary to develop adaptable and socially engaging tasks and therapy designs.

Finally, valuable comments were obtained from parents and therapists, offering suggestions for enhancing the CRI. These included the need for the robot to exhibit more personalized and active behaviours addressing children by their names, like playing an instrument, or engaging in free movement. Tozadore et al. [217] reported the significance of initial greetings, such as name recognition and high-fives to improve the quality of robot interactions be more natural. Additionally, it was noted that educational content and skills the robot can teach are crucial, and the interaction time should be managed and increased. Furthermore, some parents suggested allowing children to interact with the robot alone in the room to avoid external distractions. However, a few parents requested on playing with the robot and imitating its actions. Together, these comments highlight the requirement to reduce human involvement and maximize robot autonomy.

4.11 Concluding Remarks

We targeted to create a personalized long-term RAAT program for children with ASD and ADHD by leveraging the expertise of therapists and various behavioural measures. These measures were used to evaluate the therapy's effectiveness. We made a quantitative analysis to examine the interaction of 11 children aged 4 to 11 during RAAT. Our findings suggest that maintaining engagement in children with autism and ADHD is possible across multiple sessions, and personalizing activities based on each child's preferences leads to better engagement and attention. These results prove our hypothesis that personalized sessions are more beneficial for children with autism. However, the necessity for various robot behaviours to enhance attention and engagement among children with autism still requires more investigation. Our research clearly shows the need for more interactive, educational activities in which both children and robots participate equally. We suggested autism intervention appears to have great potential as a beneficial addition to the long-term HRI community, but additional research is required to yield more compelling results.

Our study provides valuable insights into the importance of employing customized robots in RAAT for children diagnosed with ASD and ADHD. By combining both qualitative and quantitative data, we can draw meaningful conclusions about the impact of therapy on children's behaviour as well as the possible advantages of utilizing robots in therapy. The RCRC was an ideal setting for our study due to its comprehensive rehabilitation services for children with ASD and ADHD, the long-term therapy program, personalized approach, and active parental involvement, all of which allowed us to gather valuable data and insights.

Chapter 5

Generating QAMQOR Dataset

The chapter describes a detailed overview of the process we followed to create a QAMQOR dataset, which will serve as a valuable resource for researchers studying engagement in interactions between children with autism and robots.

We begin by outlining the data collection, providing important context for our dataset. We also discuss the important steps we took during data pre-processing and feature extraction.

A significant part of our dataset creation process involved accurately labelling the collected data. We describe the techniques we employed for labelling the data, including our coding system for measuring engagement.

Overall, the QAMQOR dataset we have created provides a comprehensive resource for researchers interested in studying engagement in social interactions between children with autism and robots. Subsequent chapters of this dissertation will discuss the results of our experimental study, including how our dataset was utilized to answer key research questions.

5.1 Data Collection

Three cohorts of participants were involved in RAAT at the RCRC (center) over a one-year period from 2018 to 2019. During this RAAT study, a total of 34 children diagnosed with ASD aged between 3 and 12 years took part in the research.

Table 5.1: Characteristics of the children and information about their sessions

Child	Age	Sex	Verbal	ADHD	ADOS-2	Sessions	Mean	SD	Time
C1	5-6	F	-	-	8	6	727.73	378.15	4178
C2	5-6	M	-	-	5	6	876.53	181.62	5132
C3	8-9	M	-	-	8	6	875.33	166.27	5150
C4	3-4	M	✓	-	3	5	945.52	149.75	3978
C5	5-6	M	✓	-	4	5	760.96	184.59	3683
C6	5-6	M	✓	✓	6	5	788.28	168.04	3864
C7	7-8	M	-	-	7	4	899.45	73.73	3521
C8	3-4	M	-	-	5	2	715.2	224.8	1389
C9	5-6	M	-	-	4	2	331.6	59.6	599
C10	3-4	M	✓	-	3	2	603.6	203.6	1182
C11	8-9	M	✓	-	9	2	938.4	50.4	1821
C12	5-6	M	-	-	6	10	1150.48	166.57	11335
C13	10-11	M	✓	✓	9	9	987.11	189.57	10820
C14	7-8	M	-	✓	9	9	966.49	72.25	8447
C15	7-8	M	-	✓	8	8	1005	221.34	7950
C16	5-6	M	-	✓	8	8	963.8	215.34	7537
C17	5-6	M	-	✓	6	8	783.7	180.5	6137
C18	5-6	M	-	✓	7	8	1021.4	216.75	7996
C19	10-11	M	✓	-	9	7	1066.86	205.5	7255
C20	5-6	M	-	-	8	7	1159.66	235.34	8021
C21	12-13	M	-	-	9	6	964.67	87.83	5638
C22	8-9	M	✓	-	8	5	870.4	247.9	4352
C23	9-10	M	✓	-	8	5	977.12	162.3	4732
C24	3-4	M	-	✓	6	5	970.27	200.14	4526
C25	5-6	M	-	✓	6	4	804.2	234.41	3113
C26	3-4	M	-	✓	5	3	1270.93	335.6	3778
C27	5-6	F	-	✓	6	7	744.88	311.11	5959
C28	3-4	M	-	-	7	6	776.8	179.19	3884
C29	3-4	M	✓	-	5	7	1140.43	359.24	7983
C30	5-6	M	-	-	7	5	727	180.68	3635
C31	3-4	M	-	-	5	4	558.75	320.4	2235
C32	3-4	M	-	-	5	4	510.75	96.71	2043
C33	7-8	M	✓	-	7	6	870.5	227.59	5223
C34	3-4	M	-	-	6	2	813	14.14	1626

Of the 34 participants, 32 were boys, and only 2 were girls, reflecting the higher incidence of ASD in males. All the children had been diagnosed with ASD in combination with ADHD. The average ADOS-2 test score was 6.53 (SD = 1.74), indicating that the children in the study had moderate to severe symptoms of ASD. Twenty-three children had difficulty in communicating through speech and were labelled as nonverbal. Eleven children were able to speak and were assigned as verbal. To group the children by developmental milestones, they were divided into three age groups: from 3 to 4 years old, which had in total 10 children; from 5 to 6 years old which consisted of 13 children; and from 7 to 12 years old which included 11 children.

Table 5.1 presents further details about the participants, including the number of attended therapy sessions, and the average and total duration of those sessions. The table provides important information about the children's participation in the study, which can be used to understand their progress and to analyze the results of the interventions.

5.2 Data Pre-Processing

The datasets are valuable to the success of machine learning applications, as they provide the foundation upon which models can be built and trained [140]. In light of this, we recognized the importance of developing a robust and automated data handling pipeline to organize the raw data obtained from the RAAT study into a dataset that could be used to train and evaluate machine learning models.

To ensure that the data was well-structured and organized, we designed a data pre-processing pipeline consisting of six essential steps. This is shown in Figure 5-1.

- The initial step in our data pre-processing pipeline involved extracting frames from each video using the video capture function from the OpenCV library [28], with a rate of one frame per second.

- During the second step, we used a pre-trained "YOLOv3" model [161] from the

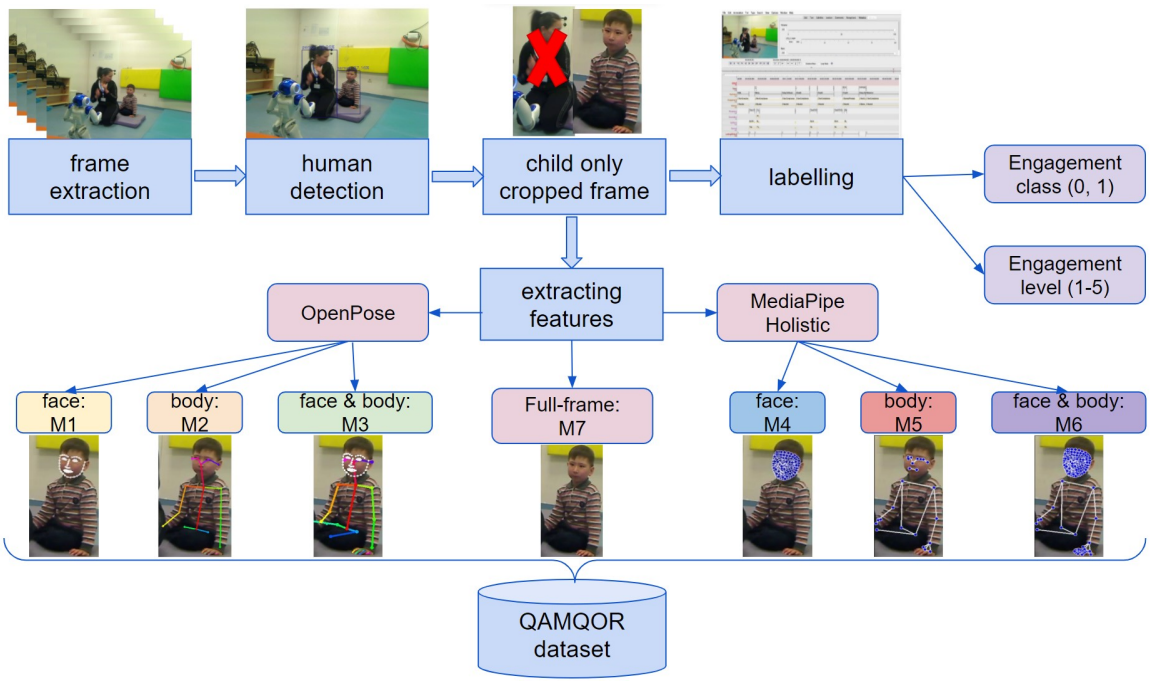

Figure 5-1: Data pre-processing steps

ImageAI library [1] to detect humans and crop the frames.

- The third step involved manually removing unnecessary images, leaving only the child's image in the frame.

- In the fourth step, we assigned labels for each frame. We used both multi-class (1-5) and binary (0 and 1) labels to capture different aspects of the child's behaviour during the therapy sessions. Further details on the labelling process are provided in Section 5.4.

- The fifth step of our pipeline involved parsing the cropped frame and saving the keypoints in a JavaScript Object Notation (JSON) file format using two libraries: the OpenPose library [32] and the MediaPipe Holistic library [89]. These libraries enabled us to extract detailed information about the child's body posture, movements, and facial expressions. More details on the feature extraction process are provided in Section 5.3.

[1] https://github.com/OlafenwaMoses/ImageAI

- Finally, we generated the dataset by extracting the facial features and body joints of the child from the JSON files into one CSV file. This file contained all the information needed to train and evaluate machine learning models, including the labels assigned to each frame.

In conclusion, our automated data handling pipeline allowed us to efficiently and effectively pre-process the raw data obtained from our study and organize it into a well-structured and informative dataset. This dataset will serve as a valuable resource for future research into ASD and machine learning applications in this field.

5.3 Feature Extraction

To extract features from the videos of children with ASD, we used the OpenPose and MediaPipe Holistic libraries. These libraries allowed us to extract both facial features, such as eye gaze, body joint movements and facial expressions. These features were chosen because they have been shown to be relevant for understanding social interaction and movement coordination difficulties in children with ASD. Facial expressions and eye gaze are important for social interaction, while body joint movements are relevant for motor coordination. The extracted features will be utilized as inputs to our models to predict diagnostic outcomes and treatment responses in children with ASD. Prior to feature extraction, we applied a Gaussian blur filter to each frame to reduce noise and enhance the quality of the features.

5.3.1 OpenPose Feature Extraction

The OpenPose library has become an essential tool for numerous research subjects that require human analysis, including human re-identification, retargeting, and HRI. This library, developed by Carnegie Mellon University, is a real-time multi-person keypoint detection tool that can estimate the body, face, hands, and foot positions [202]. With OpenPose, we can detect 2D information on 201 keypoints in the body, hands and feet and 210 keypoints in the face. OpenPose provides the 2D coordinates

of characteristic points (X and Y) and their confidence level as feature sets. Given its widespread use and effectiveness, we chose to utilise the OpenPose library [35, 36, 202] to parse the video frame in order to get keypoints of the person. Figure 5-2 shows the extracted keypoints of the face and body using different conditions.

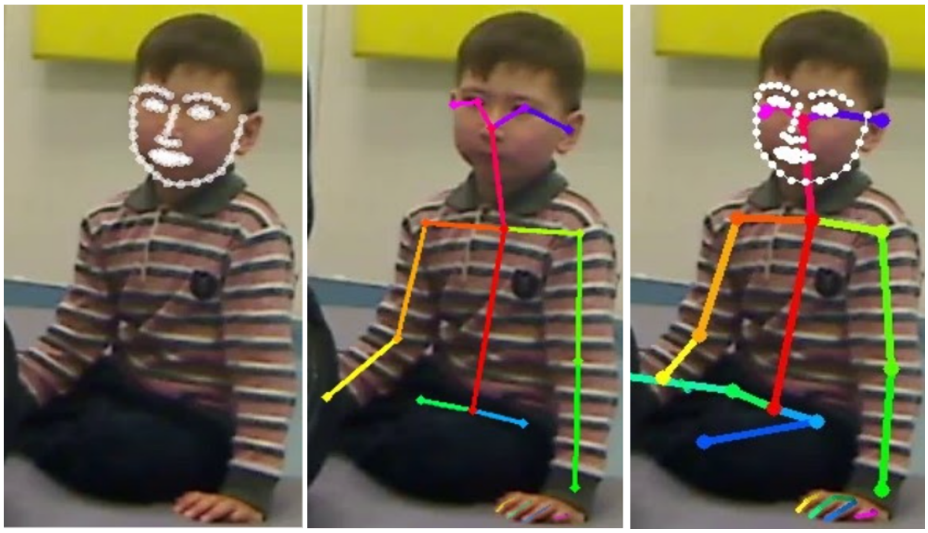

Figure 5-2: OpenPose keypoints of the child

In the end, we successfully extracted facial and body features from 93.85% of the 146,839 samples (frames) using OpenPose. However, some frames did not have an appropriately-oriented face or body, resulting in empty feature sets. To maintain the consistency of our dataset, we filled the empty features with 0s and removed them from the dataset.

5.3.2 MediaPipe Feature Extraction

To extract keypoints from the video frames, we also utilised the MediaPipe Holistic library [89]. This library is a multistage tool that integrates separate models for pose, face, and hands detection. It detects 2D information of 33 pose landmarks, 2x21 landmarks for both hands and 468 face landmarks. Unlike OpenPose, MediaPipe Holistic provides normalised X, Y coordinates, and Z-value, which provides depth information of the landmark and saves the output in JSON format. Moreover, MediaPipe Holistic indicates the likelihood of each landmark being visible in the image.

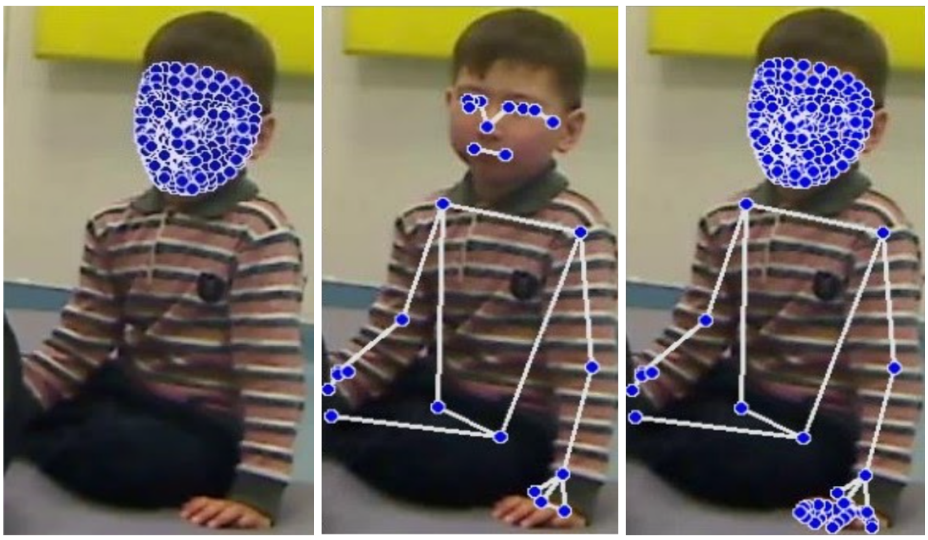

Figure 5-3: MediaPipe landmarks of the child

We chose MediaPipe Holistic to compare its performance with OpenPose and to check different image processing techniques. The face, body and hands landmarks were successfully extracted with an accuracy of 87.38% from a total of 136711 samples (frames). Figure 5-3 shows the extracted keypoints of the face and body using different conditions. Similarly, we filled the missing data in some frames with 0s and removed them from the dataset.

5.3.3 Full-frame Feature Extraction

In our study, we utilized Convolutional Neural Networks (CNNs) for full-frame feature extraction. CNNs, a subset of deep neural networks, are often employed in computer vision applications including segmentation, object identification, and picture classification. The extraction of features from pictures using CNNs has shown to be quite successful, particularly for tasks requiring local and spatial information.

We employed the EfficientNet architecture, a cutting-edge CNN architecture created to maximize the trade-off between computational efficiency and accuracy, to extract the features from the full-frame pictures. With the use of a compound coefficient, the EfficientNet technique for scaling consistently scales all depth, breadth, and resolution parameters. According to the method, a larger input picture will ne-

cessitate additional layers to expand the network's receptive area and more channels to catch the finer-grained patterns on the bigger image. The mobile inverted bottleneck convolution serves as the foundation of the EfficientNet architecture, which is a lightweight and efficient module that uses depthwise separable convolution to reduce the computational cost. EfficientNet-B7, which is the largest model in the EfficientNet family. Despite being 8.4 times smaller and 6.1 times faster at prediction than the best existing neural architecture, it obtains a top-1 accuracy of 84.3% on the ImageNet dataset.

By using EfficientNet for full-frame feature extraction, we could record and represent the significant information in the input images, which can be further utilized for the engagement recognition task.

5.3.4 Modalities

We investigated seven modalities ($M1 - M7$) in the dataset (5-1). Specifically, in the dataset, there are six modalities: $M1$ and $M4$ consist of face features, $M2$ and $M5$ consist of body and hands features, and $M3$ and $M6$ consist of all face and body features (5-1).

Additionally, we added the seventh modality ($M7$) that consists of 2048 features. These features were extracted using the pre-trained EfficientNet architecture for each sample (frame) (Section 5.3.3).

Face Features

The $M1$ and $M4$ modalities include facial features from the videos. $M1$ is a modality that consists of 210 facial features from a frame using the OpenPose tool. These features include landmarks such as the eyes, nose, and mouth, as well as facial expressions and head orientation.

On the other hand, $M4$ is composed of 1872 facial features that were extracted by utilising the MediaPipe Holistic parsing tool. This modality includes all the features from $M1$ with more detailed keypoints. The MediaPipe Holistic parsing tool is

designed to provide a more comprehensive view of the individual's facial expressions.

Both $M1$ and $M4$ modalities are used to examine the effectiveness of facial keypoints that can be used in various HRI scenarios.

Body Features

The $M2$ and $M5$ modalities include body and hand features from videos. $M2$ is a modality that consists of 201 body and hand features from a frame using the OpenPose tool. These features include key points such as the shoulders, elbows, wrists, and fingers, which allow for the recognition of body and hand movements and gestures.

On the other hand, $M5$ is composed of 300 body and hand features that were extracted by utilising the MediaPipe Holistic parsing tool. This modality includes all the features from $M2$ with more additional keypoints of the body.

Similarly, to $M1$ and $M4$ modalities, the $M2$ and $M5$ modalities can provide valuable insights into human movements and gestures, enabling advancements in a wide range of applications.

Face, Body and Hand Features

The $M3$ and $M6$ modalities include face, body and hand features from videos. $M3$ is a modality that consists of 411 face, body and hand features from a frame using the OpenPose parsing tool, which includes all the features from $M1$ and $M4$. Similarly, $M6$ is composed of 2172 face, body and hand features that were extracted by utilising the MediaPipe Holistic parsing tool, which includes all the features from $M2$ and $M5$. This modality provides a highly detailed representation of the child's body movement and an understanding of his behaviour.

5.4 Labelling

Engagement is a significant metric in the therapy of autism, as it is defined by the active and appropriate involvement in the task with the robot and/or the therapist.

In order to accurately label the dataset, we adapted a coding approach (Likert-type scale from 0 to 5) proposed in previous studies [107, 177] to ensure comparability with other researchers working in the field of engagement estimation. As ratings of 0 and 1 represented intense noncompliance and non-compliance, respectively, in our study we decided to combine label 0 (intense non-compliance) and label 1 (non-compliance) into a single category called "non-compliance." As a result, in this work, a five-point Likert scale was used. Table 5.2 describes the differences between our coding scheme and the schemes proposed by Kim et al. [107] and Rudovic et al. [177].

Table 5.2: The differences between coding schemes

#	**Our work**	**Kim et al. [107]**	**Rudovic et al. [177]**
0		Intense noncompliance – participant stood and walked away from the table on which the robot interaction took place	Evasive – child is not responding to therapist and/or NAO's prompts at all and after the prompts, or immediately, walks away from Nao
1	Non-compliance, when a child is unwilling to engage in the conversation, is not paying attention to the engagement, and is not reacting to the therapist	Noncompliance – participant hung head and refused to comply with the interviewer's request to speak to the robot	Non-compliance – child is not responding to questions or tasks by therapist (e.g., the child hung head and refused to participate in the interaction, was looking somewhere else, not paying attention to the interaction)

Table 5.2: The differences between coding schemes (cont.)

#	**Our work**	**Kim et al. [107]**	**Rudovic et al. [177]**
2	Indifference - if the therapist asks the same question more than three times before the child follows the rules	Neutral – participant complied with instructions to speak with the robot after several prompts from the confederate	Indifferent – therapist repeats the question and/or attempts the task more than 3 times until child complies with the instructions
3	Low engagement, after 2 or 3 repetitions, the child complies with the directions	Slight interest – participant required 2 or 3 prompts from the confederate before responding to the robot	Low engagement - Child complies with the instructions after 2-3 repetitions
4	Mid engagement, when a child follows directions the first time but needs assistance, such as a finger point, a name call, a demonstration of something to pay attention to, etc.	Engagement - the participant complied immediately following the confederate's request to speak to the robot or answer a question, or in which no request was made while the robot walked, and the participant maintained their gaze on the robot or looked at the confederate or robot controller without disrupting the progress of the task of speaking to the robot.	Mid engagement – a child is to the first prompt/question to perform the task but needs a bit of boost from a therapist (eg., pointing with a finger, calling by name, showing something to pay attention to and so on)

Table 5.2: The differences between coding schemes (cont.)

#	**Our work**	**Kim et al. [107]**	**Rudovic et al. [177]**
5	High engagement, when the child spontaneously interacts with the NAO robot, responds to the therapist and/or robot, and completes the necessary tasks	Intense engagement - the participant spontaneously engaged the confederate or robot (e.g., created encouraging phrases to the robot which had not been offered as examples by the confederate, or spoke to the robot spontaneously and not only when the confederate had instructed the participant to speak), or changed his or her posture (e.g., leaned forward) to non verbal interact with the robot.	High engagement – child immediately responds to the question of a therapist, following the interaction scenario and reacting with NAO spontaneously

Table 5.2: The differences between coding schemes (cont.)

Kim et al. [107] assigned engagement scores to short fixed-interval segments, i.e. one out of every four 5-second intervals. In contrast, Rudovic et al. [177] used a task-driven coding approach, where engagement scores were assigned to a video fragment starting from the task instruction until one of the engagement levels had been met. As children with autism may engage and disengage rapidly due to co-occurring ADHD, we decided to assign the engagement level to each frame (i.e. second). For example, the researchers assigned label 3 to frames 1 to 10 for that specific 10-second duration in which the child displayed behaviours associated with label 3 (low engagement).

Three independent researchers coded the frames and the agreement score was computed from the coders' pair-wise ICC, which was found to be 82.62%.

As shown in Figure 5-4, we have two types of classifications:

1. Binary class - engaged and disengaged. Engagement levels 1 and 2 were assigned to class 0, e.i. disengaged. And engagement levels from 3 to 5 were assigned to class 1, e.i. engaged;

2. Multi-class - engagement levels from 1 to 5 left as it is without any modifications.

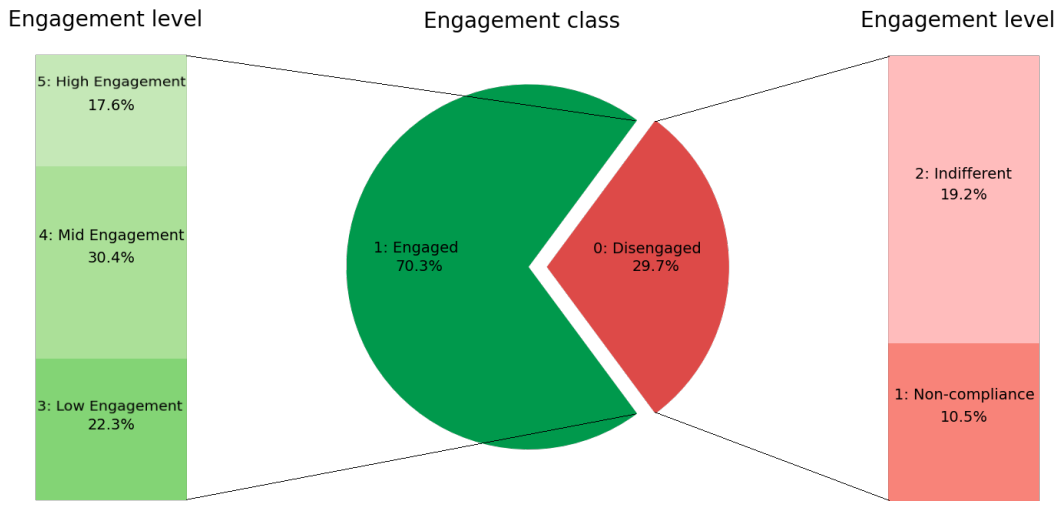

Figure 5-4: Frequency distribution of classes in the generated dataset

The frequency distribution of each label is also shown in Figure 5-4.

A lot of time and effort should be spent to assign engagement levels to each frame since the coding schema has to be trained, tested, and validated in order to produce accurate results. However, it is an important metric for evaluating robots in autism therapy. The process of analyzing behavioural data might benefit considerably from an efficient and reliable automated video coding system.

5.5 Concluding Remarks

In summary, this chapter has described the creation of the QAMQOR dataset, which consists of engagement-related data collected during 194 therapy sessions from 34 children with autism. The dataset comprises video recordings and engagement levels from more than 48 hours of video.

The QAMQOR dataset contains valuable information, including facial and body features, demographic information, and the number of children, sessions, and activities during the interaction. The dataset is a valuable resource for researchers who wish to quantitatively analyze the social behaviours of children with autism under different conditions. Additionally, the dataset can be used to train artificial models for engagement recognition during interactions with children and potentially be extended to adults with autism.

It's worth noting that combining produced QAMQOR dataset with already available HRI datasets can be a complex process, and we have to be careful about potential biases or confounding factors introduced by the combination. Therefore, the following steps were taken to be able to combine the QAMQOR dataset with other datasets:

- Data pre-processing of the QAMQOR dataset to ensure data consistency, clean any noise, and handle missing or incomplete information.
- Data alignment. The QAMQOR dataset has similar modalities of face and body features and aligned the data based on time, child and order of the sessions as in PInSoRo and DREAM datasets.
- Feature extraction, we utilized the OpenPose library to extract face and body features for having a unified set of features that can be used together.
- Ethical considerations for combining datasets involving human participants require ethical considerations. We ensure that privacy and confidentiality are protected, and data usage adheres to ethical guidelines.

Overall, this chapter has outlined the data collection and annotation process utilized for ~~data collection and~~ dataset generation. Furthermore, we discussed how we

extracted features using OpenPose and MediaPipe Holistic libraries. In subsequent chapters, we will evaluate the performance of the dataset using machine learning models and neural network architectures.

Chapter 6

Evaluating QAMQOR Dataset

Engagement recognition is a challenging problem, often addressed through supervised machine learning methods that rely on classical input features derived from spatial information, including velocity, gaze, human position, head pose, and facial expressions. In this chapter, we evaluated the QAMQOR dataset using a multimodal architecture that combines classical machine learning techniques and deep learning approaches to classify engagement.

We start by describing our methodology for evaluating the dataset in the first section. Then, in the second section, we detail the classification models used in the experiments. In the third section, we discuss the modalities present in the QAMQOR dataset employed in our evaluation. The fourth section outlined the different splits we used to partition the dataset. The fifth section described the subsets of the dataset we utilized in our experiments. We summarize the chapter with a discussion of our findings and some concluding remarks.

6.1 Methodology

To evaluate our system's performance, we employed a multimodal architecture (Figure 6-1) that utilizes classical machine learning techniques and deep learning approaches to recognize children's engagement. Given the QAMQOR dataset's noisy and missing data, we partitioned it into three subsets: training, validation, and testing. ~~and~~ More-

over, we tested the system under four different splits: Random, Child Independent, Session Independent, and Activity Independent.

We evaluated the performance of our system on two distinct classification problems: (i) binary classification (engaged and disengaged) and (ii) multi-class classification (as explained in Section 5.4). To identify the best-performing supervised machine learning models and neural network architectures, we conducted a series of experiments considering different hyperparameters (Table 6.1). For each model, we defined a search space consisting of the hyperparameters that we wanted to tune. Then, we explored this area using a search method to identify the set of hyperparameters that produced the greatest performance on a validation set.

The search algorithm used was Grid Search, where we defined a set of possible values for each hyperparameter and then trained the model using all possible combinations of these values. We chose Grid Search because it is a simple and exhaustive search method that explores all possible combinations of hyperparameters, making it a good starting point for hyperparameter tuning. However, more advanced search methods like Random Search or Bayesian Optimization could also be used to further improve the results.

Our multimodal architecture, which is shown in Figure 6-1, used three types of features: OpenPose, MediaPipe Holistic, and EfficientNet. We trained our models using different hyperparameters and reported the top-3 results of standard classification models (AdaBoost, XGBoost, and LogReg) for binary and multi-class classification problems (Table 6.2).

6.2 Classification Models

We evaluated the QAMQOR dataset using a machine learning approach that combined classical and deep learning techniques. We used several classical supervised machine learning models, such as Regularized logistic regression (LR), k-Nearest Neighbors (KNN), Extra-Trees classifier, Gaussian Naive Bayes (Gaussian NB), multivariate Bernoulli model (Bernoulli NB), Random Forest classifier, AdaBoost classifier,

Table 6.1: Supervised machine learning model types

Model	Hyperparameter	Search Space
Gaussian NB	priors	None
	var_smoothing	log-uniform distribution [1e-12, 1e-5]
Bernoulli NB	alpha	log-uniform distribution [1e-6, 10]
	binarize	uniform distribution [0.0, 1.0]
	fit_priorbool	uniform distribution [0.0, 1.0]
KNN	n_neighbors	integer uniform distribution [1, 50]
	weights	categorical distribution [‘uniform’, ‘distance’]
	algorithm	categorical distribution [‘auto’, ‘ball_tree’, ‘kd_tree’, ‘brute’]
Random Forest	n_estimators	integer uniform distribution [10, 100]
	min_samples_split	integer uniform distribution [2, 10]
	max_depth	integer uniform distribution [10, 50]
Extra Trees	n_estimators	integer uniform distribution [10, 100]
	min_samples_split	integer uniform distribution [2, 10]
	max_depth	integer uniform distribution [10, 50]
Gradient Boosting	n_estimators	integer uniform distribution [50, 200]
	max_depth	integer uniform distribution [2, 10]
	learning_rate	log-uniform distribution [0.001, 1.0]
AdaBoost	n_estimators	integer uniform distribution [10, 100]
	algorithm	categorical distribution [‘SAMME’, ‘SAMME.R’]
XGBoost	n_estimators	integer uniform distribution [50, 1000]
	max_depth	integer uniform distribution [2, 20]
LR	solver	categorical distribution [‘lbfgs’, ‘liblinear’, ‘newton-cg’, ‘newton-cholesky’, ‘sag’, ‘saga’]

Gradient Boosting classifier, and XGBoost. Moreover, for deep learning model, we used a recurrent neural network (RNN), specifically bidirectional Long Short-Term Memory (BiLSTM). We implemented these models using Scikit-learn version 0.24.2 [150], XGBoost version 1.5.0 [42], and Torch version 1.7.1 [149] in Python. To determine which models performed best, we used the hyperparameters presented in Table 6.1 and reported only the top-3 results of standard classification models (AdaBoost, XGBoost, and LogReg) for binary and multi-class classification problems.

For the sequential model architecture, we used a 2-layer BiLSTM with 256 hidden units to consider sequential dependencies. We identified three different sequences of frames, 5, 10, and 15 frames per sample, to train our models. The BiLSTM models had similar performance to the classical machine learning models. We used a

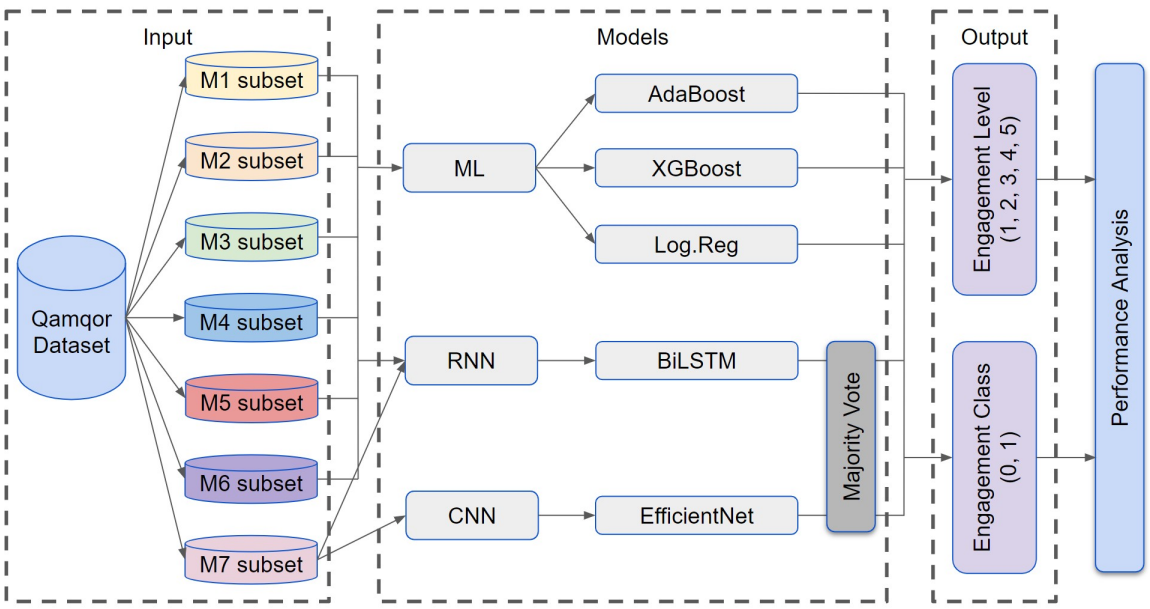

Figure 6-1: The overview of the architecture for multimodal engagement recognition

logarithmic softmax activation function for the output layer and the swish activation function for the rest of the layers. We trained our neural network models using the Adam optimizer [110] with a learning rate of 0.001 and a batch size of 60 for 20 epochs each using the cross-entropy loss function as an objective.

6.3 Results

6.3.1 Performance metrics

We have used several performance metrics to evaluate the algorithms for engagement recognition. These metrics helped to assess the effectiveness and efficiency of the models.

Accuracy is a fundamental performance metric used to evaluate engagement recognition algorithms, therefore we decided to report them in Tables 6.2, 6.3. It measured how well the algorithms correctly identified the child's level of engagement and emo-

tional response during the interaction with the NAO robot in autism therapy sessions.

F1-score metric was used to assess the algorithms' overall performance by providing a more balanced assessment as our data is highly imbalanced. Due to the space limit in the tables and the similarities of the achieved values, F1 score with accuracy values, just with smaller changes, did not record them in the table of results. XGBoost algorithm performed the best in terms of accuracy F1-Score (0.85), indicating that it correctly identifies the child's level of engagement with high precision and recall in the Random split.

Processing time was measured in seconds for each subset. It is a critical metric, especially when we will predict engagement levels in real-time, as it affects the system's responsiveness and usability. We calculated the processing time for each of the standard machine learning algorithms. Based on the result, AdaBoost has a processing time of 1.3 seconds per frame, making it relatively efficient for real-time applications, while still achieving a reasonably high accuracy.

In order to find how quickly the engagement recognition algorithms reach stable and accurate predictions, we used the speed of convergence. Smooth and steep learning curves indicated faster convergence. This convergence rate was evaluated using learning curves, which plot the performance of the algorithms over different training iterations or epochs. As a result, the LR algorithm had slower convergence, stabilizing after 100 training iterations and slightly lower accuracy and F1-Score compared to the other two algorithms, indicating that it may require more training iterations to achieve optimal performance.

6.3.2 Modalities

First, we tested H2 by analyzing the impact of different modalities on engagement recognition accuracy. To achieve this, we trained models using various combinations of modalities and analyzed their contributions. We compared six modalities from the dataset (see Section 5.3.4 for more detailed information).

To further improve the neural network model's performance, we introduced a seventh modality ($M7$) consisting of 2048 features extracted using the pre-trained

EfficientNet architecture for each sample. We trained models using actively selected samples from each modality, and possible samples were predefined to ensure dataset consistency.

By exploring different modalities, we were able to identify which features contributed the most to engagement recognition accuracy. Our findings suggest that incorporating multiple modalities can significantly improve engagement recognition accuracy.

Face Features

After analyzing the experimental results presented in Table 6.2, we found that $M1$ modality generally outperformed $M4$ modalities in terms of engagement classification accuracy, although there were a few exceptions. Specifically, in binary classification during the Session Independent split, both the XGBoost and LR algorithms achieved higher accuracy on the $M4$ modality than on the $M1$ modality. In multi-class classification during the same split, the XGBoost algorithm achieved higher accuracy on the $M4$ modality than on the $M1$ modality. Moreover, during the Activity-Independent split, AdaBoost, XGBoost, and LR algorithms achieved higher engagement classification accuracy on $M4$ modality compared to $M1$ modality in multi-class classification. These exceptions suggest that the choice of modality and algorithm may influence on the particular application and context of the engagement recognition task.

After analyzing the results presented in Table 6.2, we can conclude that the $M1$ modality consistently outperformed the other modalities in most of the splits, especially in the Child Independent and Session Independent splits. In binary classification, the $M1$ modality achieved an accuracy value of 84.17% and 77.16% in Child Independent and Session Independent splits, respectively. Similarly, in multi-class classification, the $M1$ modality achieved accuracy values of 43.21% and 39.38% in Child Independent and Session Independent splits, respectively.

Finally, we observe that having more facial features is better for the Activity-Independent split in multi-class problems to classify engagement levels for all algorithms. This is likely because children's movement varies from one activity to another,

Table 6.2: Classification accuracies of QAMQOR dataset, in %

Classes	Algorithm	OpenPose			MediaPipe Holistic			EfficientNet
		M1	M2	M3	M4	M5	M6	M7
Random Split								
Binary	AdaBoost	80.47	80.41	80.91	74.19	73.92	74.13	-
	XGBoost	84.17	**90.18**	89.78	79.60	84.25	84.41	-
	LR	80.05	77.38	80.29	73.94	74.40	74.36	-
	BiLSTM	-	-	-	-	-	-	80.59
Multi	AdaBoost	47.43	47.35	48.31	35.09	34.87	35.75	-
	XGBoost	60.70	**78.43**	77.42	51.88	65.09	64.76	-
	LR	46.73	46.38	48.66	35.52	37.18	37.65	-
	BiLSTM	-	-	-	-	-	-	57.69
Child Independent Split								
Binary	AdaBoost	**84.17**	81.36	81.36	76.48	77.04	76.58	-
	XGBoost	83.39	80.14	81.9	76.14	76.8	76.97	-
	LR	83.61	83.48	83.33	77.45	78.32	77.93	-
	BiLSTM	-	-	-	-	-	-	56.62
Multi	AdaBoost	**43.21**	38.36	39.46	29.29	28.38	27.96	-
	XGBoost	34.24	33.29	34.57	26.75	26.79	26.95	-
	LR	39.1	38.24	37.79	26.06	27.81	27.46	-
	BiLSTM	-	-	-	-	-	-	26.54
Session Independent Split								
Binary	AdaBoost	**77.16**	74.51	75.77	76.94	75.89	75.76	-
	XGBoost	75.39	72.74	73.07	75.73	76.36	76.47	-
	LR	76.31	76.36	75.97	76.67	76.32	76.41	-
	BiLSTM	-	-	-	-	-	-	68.06
Multi	AdaBoost	39.12	38.42	37.64	36.31	19.14	36.86	-
	XGBoost	35.05	33.23	34.66	35.48	34.52	35.25	-
	LR	**39.38**	38.17	37.49	37.9	37.12	36.18	-
	BiLSTM	-	-	-	-	-	-	34.36
Activity Independent Split								
Binary	AdaBoost	65.83	59.17	57.08	62.5	61.03	62.5	-
	XGBoost	68.33	66.67	71.25	62.5	66.91	66.91	-
	LR	67.08	65	64.58	62.5	66.18	67.66	-
	BiLSTM	-	-	-	-	-	-	**79.78**
Multi	AdaBoost	22.08	20.83	22.08	30.88	16.91	15.44	-
	XGBoost	16.25	26.25	30.83	29.41	36.76	30.15	-
	LR	21.67	21.25	21.25	28.68	19.85	19.12	-
	BiLSTM	-	-	-	-	-	-	**43.06**

and more features in the small area (in our case, the face) provide better accuracy. However, for the baseline (Random split) and Child Independent split, a minimal amount of facial features is sufficient.

Body and Hand Features

We conducted a thorough analysis of the experimental results presented in Table 6.2. Our analysis shows that in most cases, the $M2$ modality outperformed the $M5$ modality in terms of accuracy. However, there were some specific scenarios in which the $M5$ modality performed better, including:

- In binary classification during the Session Independent split, the AdaBoost and XGBoost algorithms achieved higher engagement classification accuracy.
- In multi-class classification during the Session Independent split, the XGBoost algorithm achieved higher engagement classification accuracy.
- In binary classification during the Activity Independent split, the AdaBoost, XGBoost, and LR algorithms achieved higher engagement classification accuracy.
- In multi-class classification during the Activity Independent split, the XGBoost algorithm achieved higher engagement classification accuracy.

Furthermore, we found that in the Random split, the $M2$ modality outperformed the other seven modalities in both binary (with a value of 90.18%) and multi-class (with values of 78.43%) classification problems using the XGBoost algorithm. Conversely, the $M5$ modality achieved the lowest engagement classification accuracy in both binary (with a value of 73.92%) and multi-class (with values of 34.87%) classification problems using the AdaBoost algorithm. Additionally, using the AdaBoost algorithm, the $M5$ modality attained the lowest accuracy of 19.14% in multi-class classification problems, which was observed in the Session Independent split.

Our results suggest that having more body and hand features is beneficial for the Activity Independent split in binary classification problems to classify engagement

levels for all algorithms. This is likely because children's movement varies from one activity to another, and more features in the small area of interest (in our case, the body and hands) provide better accuracy. However, in the baseline (Random split) and Child Independent split, a minimal amount of features is sufficient.

Face, Body and Hand Features

We examined the results from Table 6.2 and found that, in most cases, the $M3$ modality outperformed the $M6$ modality in terms of engagement classification accuracy. However, there were a few exceptions, as follows:

- The XGBoost and LR algorithms achieved higher accuracy in binary classification using the $M6$ modality compared to the $M3$ modality in the Session Independent split.
- The XGBoost algorithm achieved higher accuracy in multi-class classification using the $M6$ modality compared to the $M3$ modality in the Session Independent split.
- The AdaBoost and LR algorithms achieved higher accuracy in binary classification using the $M6$ modality compared to the $M3$ modality in the Activity Independent split.

However, despite these exceptions, the $M3$ and $M6$ modalities had the lowest overall engagement classification accuracy across all modalities. Specifically, the $M3$ modality achieved the lowest accuracy of 57.08% in binary problems with Activity Independent split, while the $M6$ modality had the lowest accuracy of 15.44% in multi-class classification problems.

However, these modalities that consist of all features achieved the lowest engagement classification accuracy among all modalities. In particular, the $M3$ modality had the lowest classification accuracy of 57.08% compared to the other seven modalities in binary problems with Activity Independent split. Moreover, the $M6$ modality achieved only 15.44% engagement classification accuracy in multi-class classification problems, which is the lowest value.

Full-frame Features

To provide a more cohesive analysis, we can expand on the results from Table 6.2. The $M7$ modality, which includes full-frame features, performed exceptionally well in Activity Independent split for both binary and multi-class classification problems. This suggests that full-frame features are more effective in this particular split, as they require more keypoints to accurately predict engagement levels. Additionally, the balanced data distribution in the Activity Independent split may have contributed to the superior performance of the $M7$ modality.

However, it is important to note that the $M7$ modality did not perform well in the Child and Session Independent splits for binary classification, obtaining the lowest engagement recognition accuracy values of 56.62% and 68.06% respectively. This could be due to the fact that the Child and Session Independent splits had more complex variations and challenges, such as different lighting conditions, backgrounds, and camera angles. Therefore, using full-frame features may not be sufficient to accurately predict engagement levels in these splits.

The results suggest that the optimal choice of modality depends on the classification problem and split used. For example, full-frame features may be more effective in Activity Independent split, while a combination of facial, body, and hand features may be more effective in Child and Session Independent splits. Understanding the strengths and weaknesses of different modalities can help improve the accuracy of engagement recognition systems in various settings.

6.3.3 Splits

We conducted experiments using four different data splits, namely Random, Child Independent, Session Independent, and Activity Independent. These splits were chosen to evaluate the performance of different machine learning algorithms in various scenarios, where data samples may have different levels of similarity. The data was divided into non-overlapping subsets, with approximately 80% used for training and 10% each for validation and testing. The algorithms were trained on a representative

sample of the data and evaluated on previously unseen data to ensure generalization.

Random Split

We randomly split the dataset into disjoint training, validation, and testing subsets as the baseline for all models. The dataset consisted of X samples and included features from Y modalities. In the following results, we compare each condition with the others.

We trained and compared several classification models using different modalities and sequence lengths. Among them, the XGBoost algorithm with $M2$ modality achieved the highest accuracy in engagement classification, with values of 90.18% and 78.43% in binary and multi-class problems, respectively. However, the AdaBoost algorithm with $M5$ modality performed poorly, achieving the lowest accuracy of 73.92% and 34.87% in binary and multi-class classifications, respectively. None of the three splits produced results that exceeded these values. In total, we trained and compared 9 models with different parameters.

Regarding computational time, the standard classification methods were faster than the neural network models. The average time for training and testing RNN models was 13 minutes. The fastest average time for the standard classification methods was 2 minutes using $M1$ modality. AdaBoost was the fastest algorithm, taking approximately 5 minutes, while XGBoost was the slowest, taking approximately 36 hours.

We also trained the BiLSTM model with three different sequence lengths of samples: 5, 10, and 15 frames per sample. The highest results were achieved using 10 frames per sample with $M7$ modality, both for binary and multi-class classification problems, with values of 80.9% and 59.37%, respectively. When using OpenPose features, we achieved the lowest accuracy when using 15 frames per second, with values of 53.69% and 45.35% in binary and multi-class classification problems, respectively.

Child Independent Split

We employed a Child Independent split to divide the videos of 34 children in training, testing, and validation subsets. This approach ensured that the training subset comprised samples from 28 children, while each testing and validation subset included samples from three children. We used non-overlapping subsets of children to train and evaluate the models to avoid overfitting.

The AdaBoost classifier outperformed other models with the highest accuracy of 84.17% in binary and 43.21% in multi-class classification problems using the $M1$ modality, which consisted of OpenPose face features (indicated by green cells in Table 6.2). However, the performance of other machine learning models varied depending on the modality and classification task. For example, the LR algorithm showed higher engagement recognition accuracy among other algorithms on $M3$ and $M6$ modalities in binary classification, but not in multi-class classification problems. The $M7$ modality showed the lowest accuracy with a value of 56.62% in binary classification problems for engagement recognition, whereas for multi-class classification problems, the LR algorithm on the $M4$ modality achieved the lowest accuracy with a value of 26.06%. The lowest results are indicated by red cells in Table 6.2.

We evaluated the BiLSTM model using facial and body features from OpenPose and MediaPipe libraries, as well as EfficientNet features. The BiLSTM model performed better with a sequence length of 15 frames per sample. As shown in Table 6.3, the highest accuracy of 58.84% was achieved in binary classification with MediaPipe features ($M6$ modality). For multi-class classification problems, the best results were obtained using EfficientNet features ($M7$ modality) with an accuracy of 44.74%. However, EfficientNet features produced the lowest accuracies among all features in the Child Independent split for both classification problems.

Session Independent Split

To analyze the impact of sessions on predicting engagement levels, we combined videos from the entire dataset to train our models. Each testing and validation subset

Table 6.3: Experimental results of RNN

Classification	Sequence	OpenPose	MediaPipe	EfficientNet
Random Split				
Binary	5	79.44	78.95	79.37
	10	79.78	79.57	**80.9**
	15	53.69	54.01	69.98
Multi-class	5	54.34	45.35	53.34
	10	52.89	52.49	**59.37**
	15	45.35	54.33	47.07
Child Independent Split				
Binary	5	56.71	57.26	53.01
	10	58.25	58.58	48.49
	15	58.04	**58.84**	53.67
Multi	5	42.24	33.39	26.31
	10	40.84	36.03	36.98
	15	37.71	33.5	**44.74**
Session Independent Split				
Binary	5	59.55	62.12	56.53
	10	62.47	**64.33**	54.02
	15	61.67	64.31	49.45
Multi	5	**44.52**	39.06	39.9
	10	42.40	38.93	42.44
	15	46.07	31.54	42.57
Activity Independent Split				
Binary	5	59.67	**64.34**	61.26
	10	61.48	63.73	54.23
	15	62.12	63.1	50.56
Multi	5	33.47	31.53	30.19
	10	33.01	32.35	31.47
	15	**34.07**	30.71	31.06

contained videos from one session not present in the training subset of the remaining session videos for each specific class. To avoid overfitting sessions, we divided the sessions without repetition into training, testing, and validation subsets.

Similar to the Child Independent split, the AdaBoost classifier achieved the highest test accuracy on all modalities in binary classification. Specifically, the best result was achieved on the $M1$ modality with an accuracy of 77.16% using OpenPose features. However, for multi-class classification, the LR algorithm showed the best performance among all modalities, with the highest accuracy of 39.38% on the $M1$ modality using OpenPose features.

Consistent with the Child Independent split, the $M7$ modality in binary classification achieved the lowest engagement recognition accuracy, with a value of 56.62%. However, in contrast to the Child Independent split, the AdaBoost algorithm showed the lowest accuracy in multi-class classification with a value of 19.14%. Other results are shown in Table 6.2.

Using the BiLSTM model, we achieved higher accuracy with a sequence length of 10 frames per sample for binary classification and 5 frames per sample for multi-class classification. The best accuracy in the multi-class classification problem was achieved on the $M3$ modality, with a value of 44.52% using OpenPose face and body features. Meanwhile, in the binary classification problem, the highest test accuracy was achieved on the $M6$ modality using MediaPipe features, with a result of 64.33%.

Activity Independent Split

To investigate the impact of activity type on predicting engagement levels, we trained our models on a combination of extracted videos from the entire dataset. To prevent overfitting, we ensured that each testing and validation subset contained videos of one activity type that were not included in the training subset for each specific class.

For the BiLSTM model, the $M7$ modality achieved the highest test accuracy among all modalities. Specifically, the best accuracy was obtained for binary classification at 79.78% and for multi-class classification at 43.06%. In contrast to the Child Independent split, the AdaBoost algorithm showed the lowest engagement recognition

accuracy with 57.08% for binary classification on the $M3$ modality and 15.44% for multi-class classification on the $M6$ modality. Other results are presented in Table 6.2.

The RNN results are presented in Table 6.3. In the BiLSTM model, we achieved the highest accuracy on the $M3$ modality using OpenPose features in the multi-class classification problem, and the best accuracy was 34.07% with 15 frames per sample. For the binary classification problem, the highest test accuracy was obtained on the $M6$ modality consisting of MediaPipe face and body features with 5 frames per sample, and the accuracy was 64.34%. In contrast to the Random split, the $M7$ modality consisting of EfficientNet features achieved the lowest accuracy in both binary and multi-class classification problems. These results are indicated by red cells in Table 6.3.

6.3.4 Subsets

Creating subsets based on child and activity from the QAMQOR dataset can help provide a more targeted view of the data and uncover patterns and trends that may not be visible in the larger dataset.

Child-based Subsets

To identify the child whose data subset outperforms other subsets, we generated 34 child-based subsets and evaluated each child using the AdaBoost classifier algorithm on face and body features extracted from OpenPose keypoints ($M3$ modality). We divided the data into training, testing, and validation sets with a ratio of 0.8-0.1-0.1, respectively. For example, if child C1 had six sessions, we randomly chose four sessions to train the model and evaluated using the other two sessions for the Session Independent split. A similar approach was applied to the Activity Independent split.

Our child-based subsets demonstrated better results than the QAMQOR dataset in the Child Independent split. Table 6.4 shows the results of engagement recognition for each child, using both Activity Independent and Session Independent splits. The

cells highlighted in red represent areas where the recognition performance was particularly poor. The cells highlighted in green represent areas where the recognition performance was particularly good.

The AdaBoost algorithm achieved the highest engagement classification accuracy on the child C33 subset compared to other children, as its values were higher in all four scenarios, with scores of 95.66% and 99.97% in binary and multi-class classification for Activity Independent split, and scores of 83.94% and 90.54% for Session Independent split. C33 had ASD only, with an ADOS-2 score of 7, but was verbal and attended six sessions.

The results for the lowest engagement classification accuracy were not consistent, with different subsets of children achieving the lowest accuracy for each binary and multi-class classification problem during the Activity and Session Independent splits. However, child C8 subset has the lowest scores in all four scenarios, with a score of 24.29% in binary classification for Activity Independent split, and scores of 5.95%, 26.14%, and 17.36% in multi-class classification for Activity Independent split, and binary and multi-class classification for session independent split, respectively. Children C8, C9, and C34 were nonverbal and diagnosed with ASD only with ADOS<=6, and their child-based subsets had only two sessions, which is a small data to train the model. In contrast, child C17 attended 8 sessions and was diagnosed with ASD and ADHD with ADOS-2 equals 6. As children diagnosed with ADHD cannot sit in one place, the model could not accurately predict their engagement levels. Almost all children who were not diagnosed with ASD and ADHD achieved higher accuracies compared to others, except for child C27.

Other interesting comparisons can be made by looking at the performance of children across the different scenarios. For example, the child C2 subset has the highest score in the multi-class classification for the Activity Independent split, but its score drops significantly in the Session Independent split, while the child C21 subset has a very consistent performance across all four scenarios.

The results of the engagement recognition experiments showed significant variability across different children and scenarios. It is important to consider individual

Table 6.4: Experimental results of child-based subsets

Subsets	**Activity**		**Session**	
	binary	multi-class	binary	multi-class
C1	65.71	39.89	58.37	29.65
C2	84.55	48.23	**84.20**	63.18
C3	78.67	48.54	71.45	29.81
C4	75.07	74.09	69.01	45.40
C5	75.98	69.88	78.37	64.22
C6	61.79	50.62	63.85	32.78
C7	71.66	54.37	43.75	18.09
C8	24.29	5.95	26.14	17.36
C9	39.29	5.72	39.17	22.42
C10	50.00	50.00	45.86	45.86
C11	46.59	33.15	47.94	24.32
C12	82.13	38.33	82.78	74.17
C13	61.73	30.58	23.11	63.27
C14	65.03	25.70	69.49	23.60
C15	53.87	39.97	36.38	10.06
C16	73.76	69.26	33.93	7.99
C17	64.07	41.88	0.44	33.29
C18	72.43	60.60	65.26	69.55
C19	**91.58**	54.85	71.31	44.81
C20	81.35	63.16	47.67	51.60
C21	77.43	44.65	64.69	75.06
C22	55.70	72.48	79.78	27.02
C23	77.11	73.07	73.64	59.52
C24	82.07	66.79	72.10	57.07
C25	70.18	46.79	47.02	35.63
C26	40.55	54.16	43.98	28.31
C27	60.44	42.07	49.00	**79.46**
C28	71.91	36.52	24.54	44.24
C29	82.98	44.75	75.57	34.02
C30	77.06	56.98	75.61	60.20
C31	65.64	83.43	66.62	39.61
C32	50.00	**88.66**	46.39	62.79
C33	**95.66**	**99.97**	**83.94**	**90.54**
C34	29.17	10.42	26.29	7.73

Table 6.5: Experimental results for the group of children from child-based subsets

Group of children	**Activity**		**Session**	
	binary	multi-class	binary	multi-class
ASD+ADHD	65.25	52.21	48.18	41.44
ASD	66.93	49.95	59.52	44.36
Verbal	70.38	**61.33**	64.76	**48.73**
Nonverbal	64.40	46.02	51.10	40.91
ADOS>6	71.60	54.29	57.11	43.93
ADOS<=6	61.07	47.41	**74.19**	42.76
3-4 years old	57.17	55.40	49.65	43.70
5-6 years old	68.75	48.17	54.43	47.82
7+ years old	**72.11**	55.24	62.91	40.88

differences and specific contexts when interpreting these results. To address this, we grouped children based on their characteristics, and the results are presented in Table 6.5.

Overall, the experiments showed that older children were better at differentiating between engaged and disengaged behaviours for each type of activity. In the binary classification problem, groups of children older than seven years achieved higher results, with 72.11% accuracy in the Activity Independent split. However, for younger children aged 3-4 years old, the AdaBoost algorithm was not able to accurately classify engagement, achieving only 57.17% accuracy.

Regarding multi-class classification problems, the results showed that verbal children achieved higher engagement classification accuracy with a value of 61.33% in the Activity Independent split, while nonverbal children achieved the lowest accuracy of 46.02%.

In summary, the engagement recognition results varied significantly depending on the children's age and communication skills, highlighting the importance of considering individual differences and specific contexts when interpreting the results.

During the Session Independent split, children diagnosed with high-functioning autism (ADOS<=6) accurately predicted engagement class in the binary problem with a value of 74.19%. However, the group of children diagnosed with ASD and ADHD achieved the lowest results among all binary classifications, with a value of

48.18%. In multi-class problems, the group of verbal children achieved higher engagement classification accuracy (48.73%) during the Session Independent split, but children aged over seven years got the least accuracy with a value of 40.88%.

Activity-based Subsets

We examined nine activity blocks of varying social mediation levels to identify which activities are most beneficial for specific subgroups of ASD. Results showed that different activities brought the attainment of all children's social objectives, but behavioural outcomes were affected by age-specific and core autism-related characteristics [240]. Section 3.4 provides details on the activities used, though we could not use all of them. We combined the "Hello&Bye" activity block with the "Action Song" block, as they are similar based on body movements.

To determine the activity during which engagement level classification could be better identified, we created activity-based subsets. Each subset included only one particular activity. We used OpenPose and MediaPipe Holistic keypoints for face and body features ($M3$ and $M6$ respectively) and applied the AdaBoost classifier algorithm to all blocks of activities. We also used the EfficientNet ($M7$) neural architecture to extract features for CNN training.

In Table 6.6, experimental findings are displayed. The cell with the highest accuracy for Child and Session Independent splits is coloured in green, and the lowest is in red. Results for binary and multi-class classification are reported. The "Imitations" activity block showed the highest engagement recognition accuracy in the Child Independent split with OpenPose and MediaPipe Holistic features, achieving 78.67% and 70.28% for binary and multi-class classification, respectively. Additionally, with the "Imitations" activity block the highest result of 75.18% and 42.26% was achieved in Session Independent split in binary and multi-class classification, respectively.

However, we could not notice the same consistency of the results during the Session Independent split. OpenPose features showed the best accuracy in engagement recognition in the "Touch Me" block activities with the value of 80.21% (this is the best result among all modalities in Session Independent Split). Although, MediaPipe

Table 6.6: Experimental results of models on activity-based dataset

	OpenPose		**MediaPipe**		**EfficientNet**	
	child	session	child	session	child	session
Activities	**Binary**					
Songs	72.17	65.36	69.73	71.12	54.75	54.48
Dances	44.48	76.55	69.5	69.34	55.08	65.09
Touch Me	53.42	**80.21**	64.65	70.85	50.76	68.14
Storytelling	54.32	61.44	58.72	**74.37**	51.86	54.14
Imitations	**78.67**	79.04	**70.28**	70.65	60.79	**75.18**
Emotions	66.15	65.85	61.81	60.51	**83.30**	56.14
Follow Me	77.94	31.67	46.59	51.32	36.28	50.81
Hello&Bye	63.54	51.91	59.46	64.42	39.13	58.44
Activities	**Multi-class**					
Songs	27.67	27.82	26.53	**33.45**	22.17	27.01
Dances	15.54	29.34	15.98	26.04	22.84	34.65
Touch Me	15.59	27.53	29.2	27.69	27.44	42.14
Storytelling	21.72	19.97	16.75	17.56	22.64	11.43
Imitations	**29.29**	**51.56**	**30.17**	25.27	26.40	**42.26**
Emotions	28.96	25.39	13.36	21.89	**43.47**	37.28
Follow Me	4.41	6.4	10.5	25.34	16	6.94
Hello&Bye	26.68	28.01	27.24	30.46	27.71	20.11

Holistic features showed the best accuracy in the "Storytelling" activity block with a value of 74.37%.

EfficientNet features achieved the highest accuracy of 83.30% and 43.47% for binary and multi-class classification, respectively, in the Child Independent split during the "Emotions" activity. For the "Songs" activity, the highest accuracy achieved was 33.45% for multi-class classification in the Session Independent Split with MediaPipe Holistic features.

The "Follow Me" activity showed the lowest accuracies (red cells in Table 6.6) in both binary and multi-class classifications, which may be due to the limited number of data. This activity had only 240 samples, with most data missing as the camera could not capture the child moving around the room holding a robot's hand.

In general, EfficientNet features achieved the best engagement recognition accuracies of 83.30% for binary classification during the Child Independent split and 51.56% for multi-class classification during the "Imitations" activity block. However, the total number of samples can affect the results. The "Imitations" activity block had the

second highest number of samples compared to others, while the "Follow Me" activity had less than 1% of the total number of samples. All results for each activity block are presented in Table 6.6.

We also used RNN models on activity-based subsets, but the results were not consistent. For the "Emotions" and "Follow Me" activity blocks, we achieved 0 accuracies of engagement recognition. However, for the "Emotions" activity block, we achieved the best accuracy of 76.49% with 10 frames per sample in binary classification. Also, the "Follow Me" activity block demonstrated the best accuracy of 87.5% in multi-class classification problems among all activities. We were unable to provide any reasonable justifications for these results. Therefore, we decided not to report them here.

6.4 Discussion

The chapter describes the results of experiments that aimed to evaluate QAMQOR dataset.

One potential explanation to a low accuracy is the noisy and missing data resulting from the different subsets of data gained from specifically sampled groups of children. This can lead to less meaningful information, affecting classification results, which include binary and multi-class classifications.

The machine learning algorithms achieve better engagement recognition accuracy using features extracted from the OpenPose library than those extracted from the MediaPipe Holistic and EfficientNet tools. The accuracy is compared across various types of splits, and the results show that the baseline Random split achieves the highest accuracy due to being trained on all children, sessions, and types of activities. However, the Child Independent split, which is trained on specific children, achieves better accuracy than the Session and Activity Independent splits.

This chapter makes two significant remarks:

- First, it presents binary (engaged or disengaged) and multi-class (five classes) classification problems for four types of splits. The dataset contains more sam-

ples of "Mid Engagement" (Figure 5-4), indicating that children with ASD were engaged more during the interaction with the robot. These children were willing to comply with the robot's instructions, but only with the help of a therapist.

- Second, the experiments show that individual Child-based and Activity-based subsets outperform the general Child Independent and Activity Independent splits, respectively. This finding suggests that the children's characteristics and social behaviour during the interaction affect the accuracy of engagement classification. The highest engagement recognition accuracy achieved for binary classification is 83.30%, and for multi-class classification, it is 51.56%, while the Activity Independent split achieves 79.78% and 43.06%, respectively.

6.4.1 Challenges and Solutions for Real-time Assessments

The following points are considered to further clarify the potential challenges and solutions related to real-time engagement assessments:

- Computational Complexity: The Adaboost algorithm performs its results in 1 minute 23 seconds, XGBOOST algorithm gets the result in 1 minute 46 seconds for testing, but it took hours for training. Therefore, the proposed approach might meet real-time processing constraints only if it will use a specific modality type on a particular split. For example, Child Independent split with M1 modality.

- Hardware: Our experiment was conducted on standard computing resources.

- Model Optimization: The potential model optimization techniques that can be applied to reduce the processing time without compromising assessment accuracy might include pruning or using lightweight architectures as future work.

- Streaming Data: The system connects to a video stream source to receive a

continuous stream of video frames. The system might set up a data buffer to temporarily store incoming video frames before they are processed.

- Real-time Processing: The system will use the AdaBoost machine learning model that is trained to recognize engagement levels based on the keypoints to classify engagement levels from the buffered video frames.

- Display Engagement Label: The system displays the engagement label in real-time by presenting it as a visual indicator, allowing a therapist to see the current engagement level as the video is being processed.

- Trade-offs: We are also should consider near-real-time engagement recognition that might be more appropriate for the RAAT scenario.

6.5 Concluding Remarks

We present our approach to evaluating the QAMQOR dataset using classical and deep learning techniques. Our multimodal architecture incorporates different types of features, and we trained both classical and deep learning models using various hyperparameters to determine the best-performing models. We compared face and body modalities and evaluated four types of splits: Random, Child Independent, Session Independent, and Activity Independent. Our experiments show that binary classification outperforms multi-class classification due to its simplicity. The highest test accuracy was achieved when evaluating OpenPose keypoints during the Child Independent split for binary classification, with values of 84.17%, and for multi-class classification, 43.21%.

Our work contributes to the development of engagement classification models for the treatment of ASD. To construct a labelled dataset, we annotated videos of RAT and applied supervised machine learning algorithms in combination with a recurrent neural network model with pre-trained EfficientNet. However, we found that the splits used to classify the data for the training, validation, and testing subsets have

an impact on how accurate the findings are. Therefore, we emphasize the importance of examining the data subsets in all splits to check their accuracy and validity, which is crucial for understanding and interpreting children's engagement levels.

Moreover, we suggest that the type of modality may affect the specific research question and available resources. For instance, if the focus is solely on facial expressions, $M1$ may be sufficient. However, if a more comprehensive understanding of the individual's body language is required, then $M4$ may be more appropriate. Additionally, the quality of input data and computational resources available for analysis should also be taken into account.

As our research involves a heterogeneous group, where participants vary significantly in terms of characteristics, experiences, and ASD conditions, it can pose challenges to the reliability of the study's results. Therefore we have used stratified sampling during the evaluation by dividing the heterogeneous group into meaningful subgroups based on relevant characteristics and analyze each subgroup separately. When sharing the obtained results of the experiments with parents, guardians and children, we adhere to the following principles:

- We use language that is both sensitive and age-appropriate, expressing our deep appreciation for their invaluable participation in the study.
- We prioritize honesty, acknowledging any limitations or potential biases in our results.
- We avoid making overly confident claims, particularly in subjective emotional measurements.

As a next step, we propose investigating transfer learning approaches to determine if existing CRI datasets can be utilized to classify engagement.

Chapter 7

A Transfer Learning Approach for Engagement Classification

This chapter aims to explore the extent to which transfer learning can aid in engagement recognition problems under different conditions. We started by presenting information regarding the datasets used in our work. We describe the characteristics of each dataset, including the number of samples, the annotation process, and any unique features that may affect model performance. Moreover, a brief discussion of the context in which the datasets were collected and how they relate to the problem of engagement recognition using keypoints was provided.

Following the dataset overview, we outline the data pre-processing tasks that we employed to prepare the data for model training. This includes tasks such as data cleaning, data normalization, feature extraction, and data augmentation. We provide a detailed explanation of the process of preparing and transforming raw data into a more suitable format and discuss how they affect the performance of the engagement recognition model.

Finally, we describe the model architecture used for model deployment and the results obtained from our experiments. We give an overview of the model's design, including the specific layers and activation functions used. We also discuss the training process, including the hyperparameters and optimization techniques used. We outline the evaluation criteria, such as accuracy, and F1 score, that was used to rate the

performance of the model. We also explore the limits of our strategy and some directions for further research. The data, model, and assessment criteria utilized in our engagement recognition tests are all thoroughly covered in this chapter.

7.1 Transfer Learning within CRI datasets

Utilizing data-driven approaches in HRI poses challenges due to the small and context-specific datasets, particularly in cases involving children with autism, where obtaining a sufficient number of manually labelled training samples can be expensive and time-consuming. However, there is a need for algorithms that can effectively learn from limited labelled training data by drawing on related labelled data or data with different distributions.

One potential solution to this challenge is transfer learning. This is a supervised machine learning approach that entails storing information learned from solving one issue and using it to solve another that is related. Unlike traditional machine learning methods, transfer learning attempts to transfer information from a prior task to a target task, enabling the re-training of a model with a different dataset and class distribution while requiring fewer data for training. Although transfer learning is commonly used in image recognition, its application in engagement classification using keypoints has not been extensively explored.

To increase the accuracy percentage of engagement label recognition for children diagnosed with ASD interacting with social robots in the QAMQOR dataset, we leveraged transfer learning techniques. Specifically, we utilized PInSoRo dataset to train our engagement recognition model using OpenPose keypoints.

We describe the processes taken to train our model using transfer learning, including the selection of the source dataset, the extraction of relevant features, and the fine-tuning of the model for the target dataset. A detailed explanation of how the model was adapted for the QAMQOR dataset, including any modifications to the architecture or hyperparameters was also provided.

7.1.1 Methodology

To enhance the accuracy of engagement label classification in the QAMQOR dataset, we explored the concept of transferring knowledge obtained from a similar dataset.

After carefully analyzing the content of various HRI datasets (as discussed in Section 2.8), we chose to utilize the PInSoRo dataset to train our initial model for engagement recognition. The decision to use the PInSoRo dataset was based on several factors. Firstly, it is publicly available, making it easily accessible for other researchers to replicate and extend our work.

The movements and facial expressions of children with autism are different from typically developed children from a medical perspective. But from a computational perspective, movement can be described by the location of joints, while facial expressions can be described by the shape changes of the location of facial features. In other words, the characteristics of the samples are the same, but the feature values may be different because both children with autism and typically developed exhibit human-like movements and expressions. Second, it provides an engagement label for each frame, which is similar to the QAMQOR dataset, allowing us to create a model that is more generalizable across different datasets. Moreover, PInSoRo stores feature extracted using the OpenPose library in a JSON file format, making the features extracted from this dataset very similar to the ones in the QAMQOR dataset.

In Section 2.8.5, we provide a more detailed overview of the PInSoRo dataset, including its characteristics and any unique features that may have an impact on the model's performance. Further, we discuss the steps taken to adapt the PInSoRo dataset for engagement recognition in the QAMQOR dataset, including any modifications made to the model architecture or hyperparameters.

Data Pre-Processing

The pre-processing of the PInSoRo dataset involved several steps to ensure the data was cleaned and prepared for analysis. First, we combined individual session CSV files into one, taking into account the different interaction conditions present in the dataset

(child-robot, child-child, and both). Therefore, we created a more comprehensive dataset that could be used for a variety of analyses. Figure 7-3 shows the steps for pre-processing the data.

Next, we removed rows with 'NAN' values for keypoints that were not recognized by OpenPose. This step was important because it allowed us to focus on the keypoint features that were shared between the PInSoRo dataset and the QAMQOR dataset. However, in doing so, several non-shared keypoints between the datasets were deleted, which is an important consideration when comparing the results of our analyses.

To label the engagement levels in the PInSoRo dataset, we used a scale of 0 to 5, similar to the QAMQOR dataset. For binary classification, we assigned levels from 0 to 2 to class 0 (disengaged), while levels from 3 to 5 were labelled to class 1 (engaged). This was done to simplify the classification process and allow for a more straightforward comparison between the two datasets.

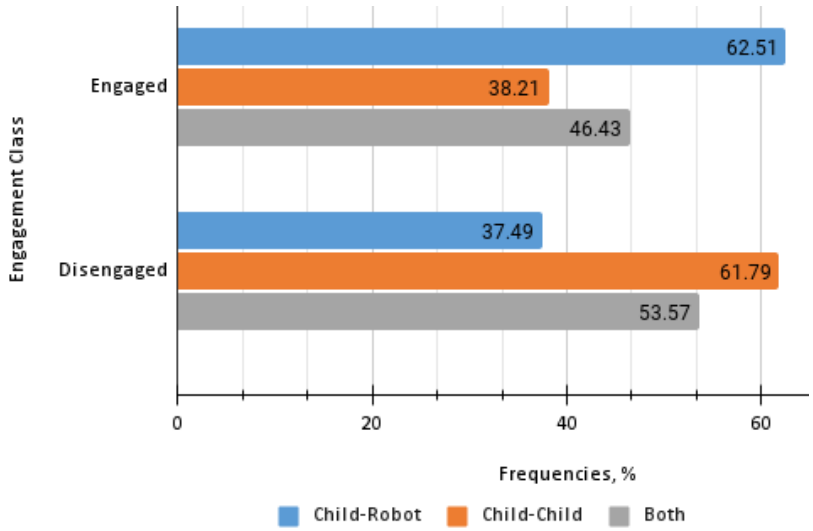

Figure 7-1: The percentage of distribution of binary engagement class in the PInSoRo dataset

The frequency distribution of classes for each interaction condition in both binary and multi-class classifications is presented in Figure 7-1 and in Figure 7-2, respectively. The QAMQOR dataset was used with OpenPose features in its original condition, which means that no pre-processing steps were taken to modify the dataset before it was used for analysis. By contrast, the PInSoRo dataset required several pre-

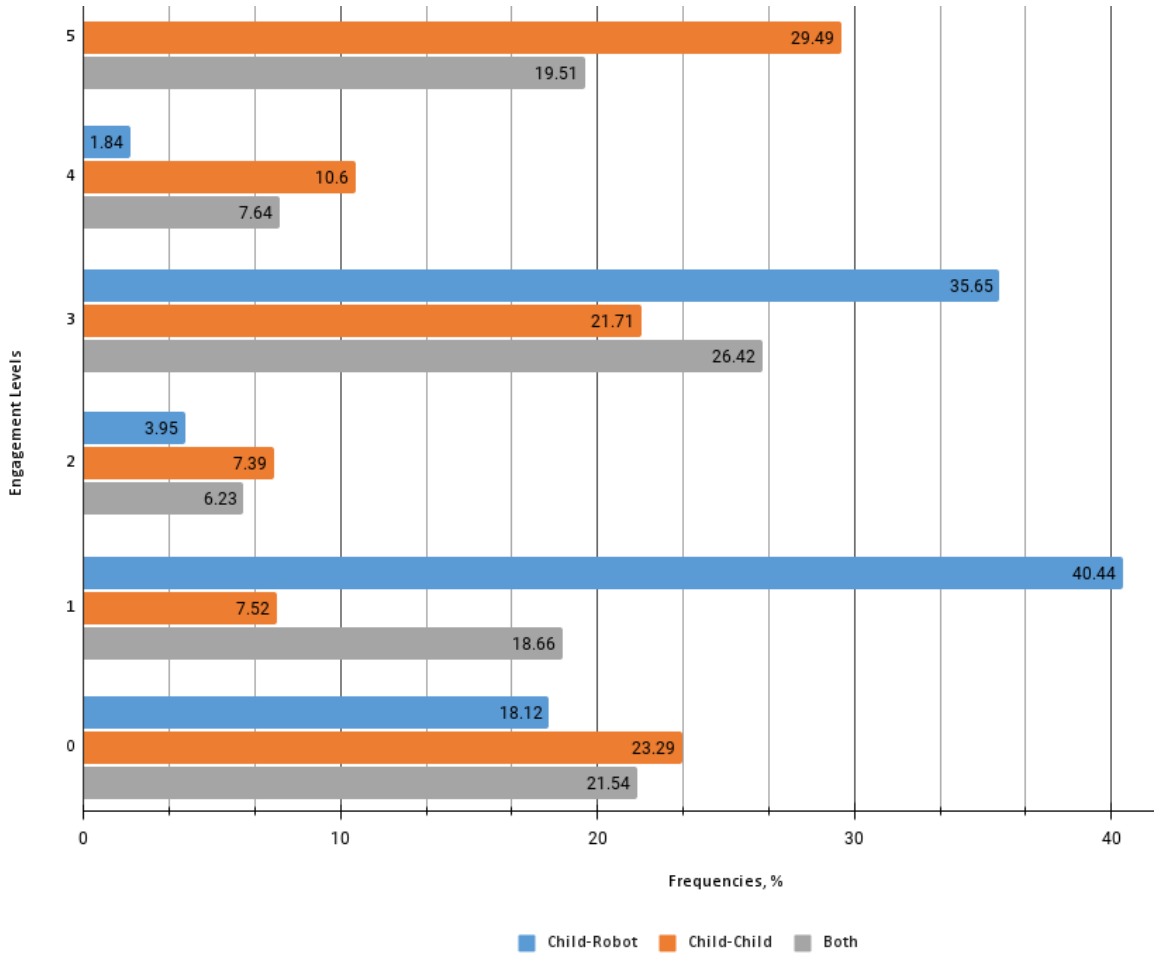

Figure 7-2: The percentage of distribution of engagement levels for multi-class classifications in the PInSoRo dataset

processing steps to clean and prepare it for analysis. Overall, these pre-processing steps were crucial to ensure that our analyses were based on accurate and reliable data.

Neural Network Architecture

Neural network (NN) architecture is a crucial component in the process of building a deep learning model. It determines the flow of information through the layers of the network and how the input data is transformed into output predictions. In this regard, there are many ways to improve and expand upon the basic architecture described above.

One approach that we used was expanding the architecture by increasing the number of hidden layers. This helped to raise the capacity of the model to identify more complex data patterns. However, it also increased the risk of overfitting the model to the training data, which lead to poor performance on new, unseen data. Therefore, we utilized regularization techniques such as dropout and weight decay to mitigate this risk.

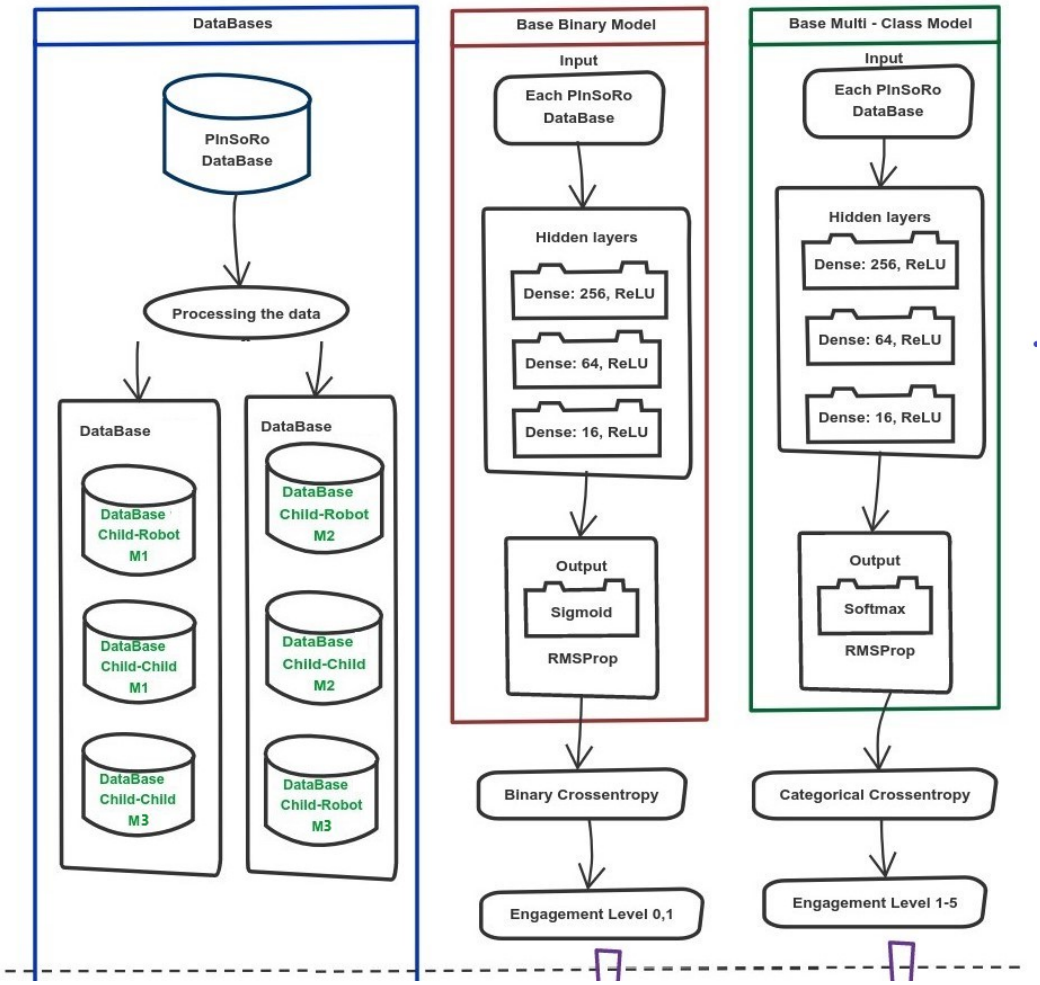

Figure 7-3: Neural Network architecture for the PInSoRo dataset

We implemented the code using open-source deep learning libraries like Keras[1] with TensorFlow[2] as the back end. We used a simple feedforward NN with three hidden layers and dropout regularization to prevent overfitting (Figure 7-3). The ReLu

[1]https://keras.io/
[2]https://www.tensorflow.org/

activation function was used for the hidden layers. A sigmoid activation function was used for a single neuron output layer. As this function is appropriate for binary classification tasks like engagement recognition. While the Softmax activation function was used for the multi-class classification. We used for binary classification the binary cross-entropy loss function. For the multi-class output, we utilized categorical cross-entropy for predicting labels. Also, an RMSprop optimizer and accuracy as the evaluation metric were used in both cases during the training. Moreover, we used the EarlyStopping callback to automatically stop training if the validation loss does not improve for two consecutive epochs. This can help prevent overfitting and save time by stopping training early if the model has already converged.

Transfer Learning

The PInSoRo datasets were used to train each multi-class and binary classification issue using three modalities ($M1$, $M2$, and $M3$) and three conditions (child-robot, child-child, and both). After pre-training the models, the output layers were frozen to prevent further training, and new models were created utilizing the pre-trained models' transferred information. The information from the PInSoRo models was then used to create the new models. The same architecture as the previously trained models were used to train these new models on the QAMQOR datasets with three modalities (Figure 7-4). Confusion matrices were plotted based on each transfer model, and the QAMQOR test set was used to assess the models' correctness.

The choice of transfer learning approach was determined based on the similarity between the target dataset and the pre-trained dataset. Freezing the layers was used when the target dataset was significantly different from the pre-trained dataset, while fine-tuning was used when the target dataset had some similarities with the pre-trained dataset.

To ensure the best possible accuracy, the performance of the transferred models was evaluated by F1 score, confusion matrices and accuracies. Additionally, different NN architectures and hyperparameters were experimented with to find the optimal combination for each specific modality within the dataset.

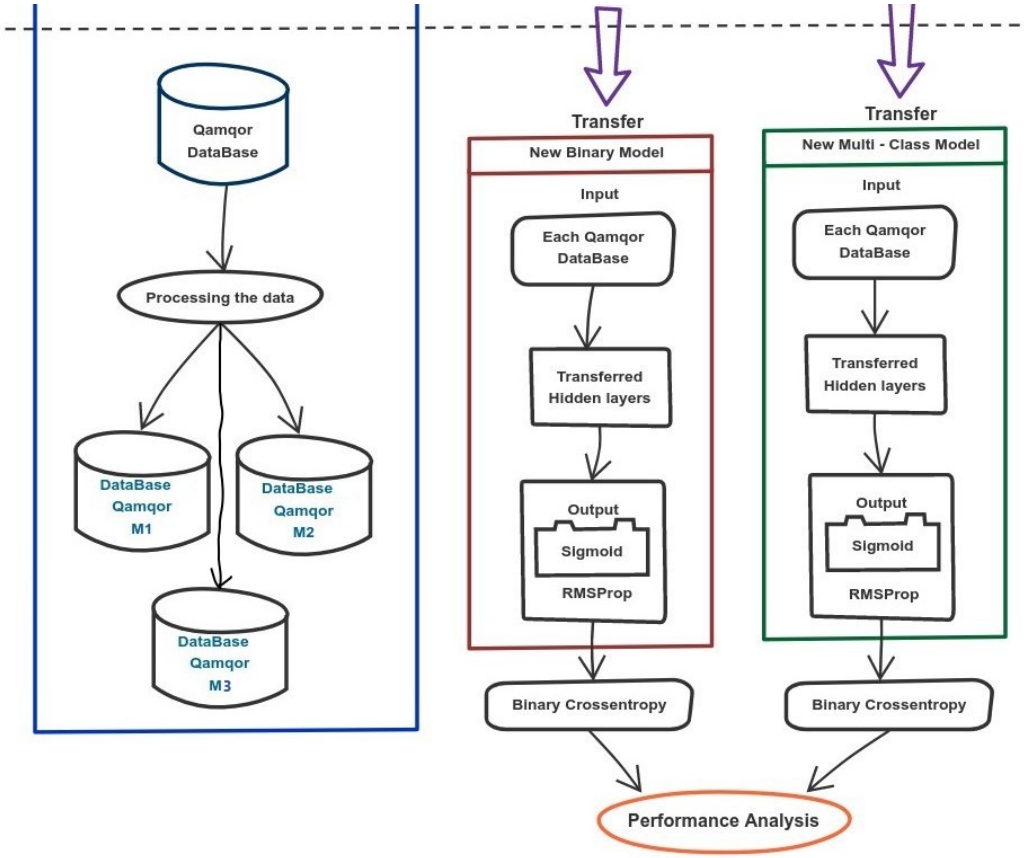

Figure 7-4: Transfer Learning approach for the QAMQOR from the PInSoRo dataset

7.1.2 Results

We conducted experiments with different classification problems and modalities. As in Section 6 modality that consists of face features showed higher accuracy of engagement recognition, we decided to examine three modalities of two datasets PInSoRo and QAMQOR:

- $M1$ which consists of face keypoints only,
- $M2$ which consists of body keypoints only,
- $M3$ which consists of both face and body keypoints.

We used the Scikit machine learning library [150] to split the datasets into training, testing, and validation set in ratios of 80%, 10% and 10%, respectively. The

experiments and analyses were run on a machine with an 8-core CPU i7 and NVIDIA GeForce RTX 2060 graphics card, which has a total memory of 16 GB.

Before applying the transfer learning approach, we evaluated the NN algorithm on various classification problems for each dataset. Table 7.1 demonstrates the results of accuracy and F1 score of engagement recognition on the PInSoRo dataset using three different modalities: $M1$, $M2$, and $M3$. The results are presented for each condition separately. Whereas, Table 7.2 shows the experimental results of QAMQOR dataset.

Table 7.1: Experimental results of accuracy and F1 score of engagement recognition on PInSoRo dataset

Conditions	**M1 modality**		**M2 modality**		**M3 modality**	
	ACC, %	**F1, %**	**ACC, %**	**F1, %**	**ACC, %**	**F1, %**
Binary classification						
Child-Robot	65.77	64.87	78.44	64.2	73.27	61.4
Child-Child	64.87	76.98	66.39	75.79	65.48	77.78
Both	61.63	69.98	61.79	71.7	64.77	65.89
Multi-class classification						
Child-Robot	52.68	35.64	55.47	42.97	56.24	43.01
Child-Child	37.96	8.49	43.61	16.74	43.25	12.07
Both	36.92	5.92	43.01	13.6	43.57	16.51

We used the PInSoRo dataset to perform transfer learning under a variety of scenarios, including child-robot, child-child, and a mix of both. We evaluated engagement recognition classifiers that were binary and multi-class. We decided to report accuracy together with an F1 score, as our data is highly imbalanced. And the F1 score provides a better performance assessment of the model with such data. We should note that the higher F1 score the better model is.

Child-Robot Condition

The results for the Child-Robot condition show that using the $M2$ modality, which includes only body keypoints, achieved the highest accuracy of 78.44% compared to the $M1$ modality, which includes only face keypoints, and the $M3$ modality, which includes both face and body keypoints. The accuracy for the $M1$ modality is 65.77%

and for M3 it is 73.27%. However, the F1 score for M1 is the highest (64.87%) among the three modalities, followed by M2 (64.2%) and M3 (61.4%). This suggests that for engagement recognition in the Child-Robot condition, body movements may be more important than facial expressions.

On the other hand, the F1 score for the $M3$ modality was the lowest among the three modalities, which may indicate that facial expressions together with body pose are not sufficient for accurately recognizing engagement in this condition.

In the multi-class classification problem, the model performs better when using the $M3$ modality compared to $M1$ or $M2$. The accuracy for $M3$ is 56.24%, while for M1 it is 52.68% and for M3 it is 55.47%, which means that the model is correct in its prediction slightly more than half of the time. Similarly, the F1 score for M3 is the highest (43.01%) among the three modalities, followed by $M2$ (42.97%) and M1 (35.64%). Overall, for the multi-class classification problem, the accuracy and F1 score are lower compared to the binary classification problem, as expected.

Overall, these results suggest that a combination of both facial expressions and body movements is important for recognizing engagement in the Child-Robot condition, and that body movement may be more important for achieving high accuracy in this task.

Child-Child Condition

In the binary classification problem, the Child-Child condition the accuracy of engagement recognition achieved a value of 64.87% and the F1 score of 76.98% using $M1$ modality for face keypoints. While using the $M2$ modality for body keypoints, the accuracy increased to 66.39% with an F1 score of 75.79%, which is a relatively small improvement over the M1 modality. Combining both face and body keypoints (M3 modality) resulted in accuracy of 65.48% and F1 score of 77.78%, which indicates a better balance between precision and recall. This result is the best performance in the binary classification task among all modalities.

In the multi-class classification problem, the accuracy of engagement recognition achieved a value of 37.96% and F1 score of 8.49% using $M1$ modality for face key-

points. However, using the $M2$ modality for body keypoints, the accuracy improved to 43.61% with an F1 score of 16.74%, which is a relatively significant improvement over the M1 modality. $M3$ modality achieved accuracy of 43.25% and an F1 score of 12.07%, which is a slightly lower accuracy than the $M2$ modality.

Overall, for the child-child conditions, using the $M2$ modality for body keypoints is the best choice for recognizing engagement with higher accuracy and F1 score in both binary and multi-class classification tasks. However, when both face and body keypoints are combined, the performance is still better than using only face keypoints (M1 modality) for the binary classification task, but not as good as $M2$ modality for the multi-class classification task.

Combination of Both Conditions

In binary classification, the accuracy and F1 scores were observed to be lower for the "Both" condition compared to the "Child-Robot" and "Child-Child" conditions in all three modalities. These results suggest that engagement recognition for the "Both" condition is more challenging than the other two conditions. Among the three modalities, the $M1$ modality performed the worst, achieving the lowest accuracy with a value of 61.63% and an F1 score of 69.98%.

In the multi-class classification task, the $M1$ modality once again achieved the lowest accuracy among all three modalities and conditions, with a value of 36.92%, and the F1 score was found to be 5.92%. The accuracy improved slightly during the $M2$ modality, achieving 43.01%, while the F1 score was found to be 13.6%. Furthermore, the accuracy increased among the three modalities in the "Both" conditions with values of 43.57% and an F1 score of 16.51%. However, these results indicate that the model's overall performance was relatively poor.

It is necessary to highlight that the model's performance achieved generally lower for the "Both" condition in comparison with the other two conditions (Child-Robot and Child-Child). This observation suggests that recognizing engagement levels in communications between both a child and a robot, and between two children, may be more challenging than recognizing engagement levels in interactions between a child

and a robot or between two children separately.

QAMQOR after Transfer Learning

Table 7.2: Experimental results of accuracy and F1 score of engagement recognition on QAMQOR dataset

Classes	M1 modality		M2 modality		M3 modality	
	ACC, %	F1, %	ACC, %	F1, %	ACC, %	F1, %
Binary	73.13	61.78	73.12	61.78	73.13	61.78
Multi-class	32.69	0	34.91	0.28	33.67	0

Our experiments applying transfer learning from the PInSoRo dataset did not yield improved performance compared to our baseline model. The accuracy and F1 score remained the same across all three modalities and conditions. These results suggest that the PInSoRo dataset may not be directly transferable to our target dataset for engagement recognition, or that the architecture of our model did not allow for the effective transfer of knowledge. Further investigation may be necessary to determine the cause of these results.

7.1.3 Discussion

According to the results of our experiments, the recognition of engagement levels in children with autism and typically developed children is a challenging task. It is feasible through the use of computer vision techniques. Our findings demonstrate that the accuracy and F1 scores for engagement recognition were lower for the "Both" condition in comparison with the "Child-Robot" and "Child-Child" conditions, indicating that the engagement recognition in interactions between a child and a robot and between two children is more challenging than in interactions between a child and a robot or between two children separately.

Our results also revealed that the performance of the model for "Both" condition was generally lower in comparison with the other two conditions, suggesting that it

could be more challenging to recognize engagement levels in sessions during interaction of both a child and a robot and between two children, than in interactions between a child and a robot or between two children separately. However, our findings suggest that transfer learning approaches may not necessarily result in improved performance in this particular task.

It is worth noting that while the movements and expressions of children with autism may differ from typically developed children, our approach using positional displacement of joints and facial features was able to effectively capture these differences and recognize engagement levels in both groups. Our results highlight the importance of developing accurate and effective algorithms for recognizing engagement levels in children with autism, as this can have significant implications for their social and emotional development. Further research is required to improve the accuracy and reliability of the engagement recognition model and to better understand the differences in engagement levels between children with autism and typically developed children.

7.1.4 Concluding Remarks

We examined H3 hypothesis that states, utilizing transfer learning from a CRI dataset (PInSoRo dataset) would enhance the engagement recognition accuracy for the QAMQOR dataset.

Unfortunately, after applying the transfer learning approach, we did not observe any improvement in the engagement recognition performance in the QAMQOR dataset. The accuracy (73.13%) and F1 score (61.78%) remained the same as before applying the transfer learning technique. This suggests that the features learned from the PInSoRo dataset were not directly applicable to our dataset or were not sufficient to improve the performance.

While transfer learning did not increase the accuracy of the QAMQOR dataset, we cannot conclude that it is not effective due to its limitations. The primary contribution of this section is in establishing a methodology for researching the formulated problem, which can guide future research in this area. As such, we have proposed

several recommendations for future work:

- To explore other available HRI datasets such as the DREAM and MHHRI datasets. The DREAM dataset may provide more definitive outcomes as it was gathered from children with autism like the QAMQOR dataset. It is recommended that the two datasets used for transfer learning should share as many features as possible for optimal results.

- More NN architectures can be tried, varying the combination of hidden layers and the nodes number in each layer. The method could benefit from using Deep Learning and RNN.

Creating a baseline network allowed us to start every transfer learning with identical conditions, making comparisons simple. Every transfer learning was performed with 5-fold cross-validation for the training and validation data. The transfer learning ran for 50 epochs, weights were saved on each epoch and we could then evaluate whichever test set we required with the weights that had the lowest validation loss.

7.2 Transfer Learning within Activity-based subsets

In the previous section, we tested our hypothesis that transfer learning from a similar dataset would lead to improved accuracy in recognizing engagement levels for the QAMQOR dataset. However, our experiment showed that it failed, as there was no improvement in engagement recognition accuracy for both binary and multi-class classification problems (see Section 7.1.2).

Nonetheless, there is another transfer learning approach that is worth exploring: transferring between datasets with similar features. To this end, we will leverage the concept of transfer learning to fine-tune a model trained on one activity for classification in the other activity data within the QAMQOR dataset.

7.2.1 Methodology

Data

This section describes the data used in our transfer learning experiments. The QAMQOR dataset comprises eight activities that were split by ID: "Songs," "Dances," "Touch Me," "Follow Me," "Imitation," "Storytelling," "Emotions," and "Hello&Bye." Each subset was evaluated separately to assess its engagement level. We excluded the "Social Acts" subset from our transfer learning experiments. Because it was introduced only in the last cohort of participants and not present in all data.

To evaluate the engagement level of children during each activity, we extracted OpenPose features to obtain keypoints of the child's face, body, and hands. These features were used to train engagement recognition models, as described in Chapter 6. The consistency in the distribution of OpenPose features across activity-based subsets in the QAMQOR dataset is crucial for transfer learning experiments since pre-trained models are leveraged to improve the performance of a new task. Having similar data distributions helps ensure that the pre-trained models transfer well to the new task.

Neural Network Model

The Keras Sequential API was used to implement NN model with several layers of dense neurons, dropout regularization, and activation functions in the output layer. This parameter plays a crucial role in the size of the weight matrix along with the bias vector. We used the non-linear Rectified Linear Unit (ReLU) function for the activation parameter, which outputs the input directly if it is positive, and zero otherwise. Moreover, it adds more sensitivity to the weighted sum. The model architecture consists of 7 hidden layers, each with a different number of neurons, and an output layer with 6 neurons similar to the number of classes in the classification problem. Similarly to the previous experiment, the Adam optimizer is used for training the model with the categorical cross-entropy loss function, for multi-class classification problems. And binary cross-entropy is used for the loss function, for binary classifica-

tion problems. The two metrics, F1-score and accuracy are used to assess the model during training and testing.

The dropout layers are used to avoid overfitting by driving the surviving neurons to acquire more robust features by randomly removing a portion of the neurons during training. We set the number of epochs to 50, the learning rate to 0.01, the chunk length to 200, and the batch size to 1000 while varying the hidden size and layer count.

Experiments

In order to conduct the experiment the following procedures were performed:

1. QAMQOR dataset was split based on the activity ID, resulting in eight different subsets: Songs, Dances, Touch Me, Follow Me, Imitation, Storytelling, Emotions and Hello&Bye.

2. Each activity subset was saved as a separate CSV file for further processing.

3. A NN model was created to further transfer the learning approach as the backbone architecture.

4. One activity subset was selected to fine-tune the pre-trained model. The model was trained on this subset using an 80/20 training/validation split.

5. The learned weights from the domain activity were used to initialize a new model for each of the remaining activity subsets.

6. The process was repeated five times per activity subset for cross-validation.

7. Steps from 4 to 6 repeated for all remaining subsets of activities.

This detailed methodology provides a clearer understanding of the experimental procedures and can be used to replicate the experiment.

7.2.2 Results

After all five folds have been processed, the mean and standard deviation of the validation accuracies and F1 scores are printed to assess the performance of the model using cross-validation. To enhance clarity and precision, we have chosen to present our findings by providing results for each activity subset separately. The cells in green indicate an improvement in the accuracy and F1 score after applying transfer learning, while the cells in red indicate a decrease in accuracy and F1 score.

"Songs" Activity

Table 7.3 presents the achieved results of engagement recognition accuracy and F1 score of the model on the "Songs" activity subset before and after applying the transfer learning approach. During binary classification the "Songs" activity subset achieved 72.24% accuracy and 83.88% F1-score. When the model was pre-trained on the "Touch Me" activity and then fine-tuned on the "Songs" activity subset, we achieved the highest improved accuracy and F1 score with values of 74.81% and 85.56%, respectively. As can be seen from Table 7.3, the results show that applying transfer learning has improved the accuracy and F1 score of the model pre-trained on the four activities ("Dances," "Touch Me," "Imitation" and "Emotions.").

On the other hand, we noted that the accuracy and F1 score for multi-class classification remained the same while pre-trained on activities, such as "Touch Me," "Imitation," "Emotions" and "Hello&Bye." When the model pre-trained on the "Dances" activity, the F1 score was slightly improved compared to the values that were achieved without applying transfer learning.

"Dances" Activity

Table 7.4 shows the achieved results of engagement recognition accuracy and F1 score of the model on the "Dances" activity subset before and after applying the transfer learning approach. During binary classification the "Dances" activity subset achieved 74.4% accuracy and 85.32% F1-score. When the model was pre-trained

Table 7.3: The results of accuracy and F1 score before and after applying the transfer learning approach for the "Songs" activity subset.

	Binary		**Multi-class**	
	Before transfer learning			
Target Activity	ACC, %	F1, %	ACC, %	F1, %
Songs	**72.24**	**83.88**	**32.51**	**9.81**
Domain Activity	After transfer learning			
Dances	73.32	84.6	32.51	9.82
Touch Me	**74.81**	**85.56**	32.51	9.81
Storytelling	69.63	82.06	23.62	7.64
Imitation	73.9	84.98	32.51	9.81
Emotions	74.45	85.33	32.51	9.81
Follow Me	64.17	76.55	31.08	11.6
Hello&Bye	68.49	81.23	32.51	9.81

on the "Touch Me" activity and then fine-tuned on the "Dances" activity subset, we achieved the highest improved accuracy and F1 score with values of 74.67% and 85.48%, respectively. As can be seen from Table 7.4, the results show that applying transfer learning has improved the accuracy and F1 score of the model pre-trained on the two activities ("Touch Me" and "Follow Me.").

Table 7.4: The results of accuracy and F1 score before and after applying the transfer learning approach for the "Dances" activity subset.

	Binary		**Multi-class**	
	Before transfer learning			
Target Activity	ACC, %	F1, %	ACC, %	F1, %
Dances	**74.4**	**85.32**	**33.46**	**10.04**
Domain Activity	After transfer learning			
Songs	73.32	84.6	33.45	10.02
Touch Me	**74.67**	**85.48**	33.46	10.02
Storytelling	71.22	83.15	18.96	6.3
Imitation	74.07	85.09	33.46	10.02
Emotions	74.43	85.33	33.46	10.02
Follow Me	66.3	78.5	33.31	13.3
Hello&Bye	70.46	82.59	33.46	10.02

On the other hand, we noted that the accuracy and F1 score for multi-class classification remained the same while pre-trained on activities, such as "Touch Me,"

"Imitation," "Emotions" and "Hello&Bye." However, when the model pre-trained on the "Storytelling" activity, the accuracy and F1 score were dramatically decreased compared to the values that were achieved without applying transfer learning.

"Touch Me" Activity

Table 7.5: The results of accuracy and F1 score before and after applying the transfer learning approach for the "Touch Me" activity subset.

	Binary		**Multi-class**	
	Before transfer learning			
Target Activity	ACC, %	F1, %	ACC, %	F1, %
Touch Me	**77.38**	**87.25**	**32.25**	**9.75**
Domain Activities	After transfer learning			
Songs	74.67	85.48	32.25	9.75
Dances	74.67	85.48	32.31	9.86
Storytelling	72.76	84.17	23.72	7.67
Imitation	74.89	85.63	32.25	9.75
Emotions	75.17	85.81	32.25	9.75
Follow Me	68.57	80.32	**32.75**	**11.76**
Hello&Bye	72.19	83.76	32.25	9.77

Table 7.5 shows the achieved results of engagement recognition accuracy and F1 score of the model on the "Touch Me" activity subset before and after applying the transfer learning approach. During binary classification the "Touch Me" activity subset achieved 77.38% accuracy and 87.25% F1-score. In contrast to the previous two activities, we did not notice any improvements, only a decrease in all values. And the lowest accuracy was achieved while the model was pre-trained on the "Follow Me" activity subset.

On the other hand, we noted improvements in the accuracy and F1 score for multi-class classification while pre-trained on activities, such as "Dances," "Follow Me" and "Hello&Bye." When the model was pre-trained on the "Follow Me" activity and then fine-tuned on the "Touch Me" activity subset, we achieved the highest improved accuracy and F1 score with values of 32.75% and 11.76%, respectively.

"Storytelling" Activity

Table 7.6: The results of accuracy and F1 score before and after applying the transfer learning approach for the "Storytelling" activity subset.

	Binary		**Multi-class**	
	Before transfer learning			
Target Activity	ACC, %	F1, %	ACC, %	F1, %
Storytelling	**67.01**	**80.25**	**23.91**	**7.73**
Domain Activity	After transfer learning			
Songs	72.76	84.17	23.24	7.54
Dances	72.76	84.17	23.24	7.56
Touch Me	72.76	84.17	23.24	7.54
Imitation	73.32	84.55	23.24	7.54
Emotions	**73.54**	**84.69**	23.24	7.54
Follow Me	68	80.08	**22.26**	**8.86**
Hello&Bye	71.15	83.05	23.25	7.56

Table 7.6 shows the achieved results of engagement recognition accuracy and F1 score of the model on the "Storytelling" activity subset before and after applying the transfer learning approach. During binary classification the "Storytelling" activity subset achieved 67.01% accuracy and 80.25% F1-score. As can be seen from Table 7.6, we noted improvements in the accuracy and F1 score for binary classification while pre-trained on all activities. And, when the model was pre-trained on the "Emotions," we achieved the highest improved accuracy and F1 score with values of 73.54% and 84.69%, respectively.

On the other hand, we noted that the accuracy and F1 score for multi-class classification slightly decreased while pre-trained on all activities compared to the values that were achieved without applying transfer learning. And when the model pre-trained on the "Follow Me" activity, the result showed the lowest values of accuracy with values of 22.26%. In contrast, the F1 score increased from a value of 7.73% to 8.86%.

Table 7.7: The results of accuracy and F1 score before and after applying the transfer learning approach for the "Imitation" activity subset.

	Binary		**Multi-class**	
	Before transfer learning			
Target Activity	ACC, %	F1, %	ACC, %	F1, %
Imitation	**75.55**	**86.07**	**34.75**	**10.31**
Domain Activity	After transfer learning			
Songs	73.32	84.55	34.75	10.32
Dances	73.32	84.55	34.75	10.32
Touch Me	73.32	84.55	34.75	10.31
Storytelling	73.32	84.55	25.85	8.22
Emotions	73.87	84.92	34.75	10.31
Follow Me	69.03	80.89	32.35	12.2
Hello&Bye	71.89	83.56	34.74	10.32

"Imitation" Activity

Table 7.7 shows the achieved results of engagement recognition accuracy and F1 score of the model on the "Imitation" activity subset before and after applying the transfer learning approach. During binary classification the "Imitation" activity subset achieved 75.55% accuracy and 86.07% F1-score. Similar to the "Touch Me" activity, we did not notice any improvements, only a decrease in all values. And the lowest accuracy was achieved while the model was pre-trained on the "Follow Me" activity subset.

On the other hand, we noted that the accuracy for multi-class classification remained the same while pre-trained on activities, such as "Songs," "Dances," "Touch Me" and "Emotions." However, when the model pre-trained on the "Storytelling" activity, the accuracy and F1 score were dramatically decreased compared to the values that were achieved without applying transfer learning.

"Emotions" Activity

Table 7.8 shows the achieved results of engagement recognition accuracy and F1 score of the model on the "Imitation" activity subset before and after applying the transfer learning approach. During binary classification the "Emotions" activity subset

achieved 76.66% accuracy and 86.78% F1-score. Similar to the "Touch Me" and "Imitation" activities, we did not notice any improvements, only a decrease in all values. And the lowest accuracy was achieved while the model was pre-trained on the "Follow Me" activity subset.

Table 7.8: The results of accuracy and F1 score before and after applying the transfer learning approach for the "Emotions" activity subset.

	Binary		**Multi-class**	
	Before transfer learning			
Target Activity	ACC, %	F1, %	ACC, %	F1, %
Emotions	**76.66**	**86.78**	**42.65**	**11.96**
Domain Activity	After transfer learning			
Songs	73.87	84.92	42.65	11.96
Dances	73.87	84.92	42.65	11.96
Touch Me	73.87	84.92	42.65	11.96
Storytelling	73.87	84.92	19.12	6.42
Imitation	73.87	84.92	42.65	11.96
Follow Me	69.91	81.57	38.49	13.85
Hello&Bye	72.57	84.02	42.66	11.98

On the other hand, we noted that the accuracy and F1 score for multi-class classification remained the same while pre-trained on activities, such as "Songs," "Dances," "Touch Me," and "Imitation." And, when the model pre-trained on the "Storytelling" activity, the accuracy and F1 score were dramatically decreased compared to the values that were achieved without applying transfer learning. However, when the model pre-trained on the "Hello&Bye" activity, the accuracy (42.66%) and F1 score (11.98%) were slightly increased compared to the values that were achieved without applying transfer learning.

"Follow Me" Activity

Table 7.9 shows the achieved results of engagement recognition accuracy and F1 score of the model on the "Follow Me" activity subset before and after applying the transfer learning approach. During binary classification, the "Follow Me" activity subset achieved 57.92% accuracy and 70.71% F1-score.

Table 7.9: The results of accuracy and F1 score before and after applying the transfer learning approach for the "Follow Me" activity subset.

	Binary		**Multi-class**	
	Before transfer learning			
Target Activity	ACC, %	F1, %	ACC, %	F1, %
Follow Me	**57.92**	**70.71**	**32.08**	**17.6**
Domain Activity	After transfer learning			
Songs	72.79	84.17	23.33	8.96
Dances	72.79	84.17	23.33	8.96
Touch Me	72.79	84.17	23.33	8.96
Storytelling	72.79	84.17	3.75	1.51
Imitation	72.79	84.17	23.33	8.96
Emotions	72.79	84.17	23.33	8.96
Hello&Bye	71.78	83.47	23.33	8.96

As can be seen from Table 7.9, we noted the highest improvement in accuracy and F1 score (with values of 72.79% and 84.17%, respectively) for binary classification while pre-trained on all activities, except for "Hello&Bye."

On the other hand, we see from Table 7.9 that the accuracy and F1 score for multi-class classification decreased while pre-trained on all activities. Also, when the model pre-trained on the "Storytelling" activity, the accuracy and F1 score were dramatically decreased compared to the values that were achieved without applying transfer learning.

"Hello&Bye" Activity

Table 7.10 shows the achieved results of engagement recognition accuracy and F1 score of the model on the "Hello&Bye" activity subset before and after applying the transfer learning approach. Similar to the "Storytelling" and the "Follow Me" activities, during binary classification, the "Hello&Bye" activity subset achieved 64.73% accuracy and 78.58% F1-score. As can be seen from Table 7.10, we noted the highest improvement in accuracy and F1 score (with values of 71.78% and 83.47%, respectively) for binary classification while pre-trained on all activities, except for "Follow Me."

Moreover, for multi-class classification, we noted that the accuracy and F1 score

Table 7.10: The results of accuracy and F1 score before and after applying the transfer learning approach for the "Hello&Bye" activity subset.

	Binary		**Multi-class**	
	Before transfer learning			
Target Activity	ACC, %	F1, %	ACC, %	F1, %
Hello&Bye	**64.73**	**78.58**	**26.05**	**8.3**
Domain Activity	After transfer learning			
Songs	71.78	83.47	**27.12**	**8.57**
Dances	71.78	83.47	27.1	8.52
Touch Me	71.78	83.47	27.1	8.52
Storytelling	71.78	83.47	24.51	7.87
Imitation	71.78	83.47	27.1	8.52
Emotions	71.78	83.47	27.1	8.52
Follow Me	69.06	80.96	25.73	10.05

for multi-class classification slightly increased while pre-trained on activities, such as "Songs," "Dances," "Touch Me," "Imitation" and "Emotions." However, when the model pre-trained on the "Storytelling" the accuracy and F1 score were dramatically decreased compared to the values that were achieved without applying transfer learning.

7.2.3 Discussion

Transfer learning results were highly volatile and required further investigation before reporting. The number of folds was limited to 5 to train per transfer learned model.

According to the results, we evaluated the performance of their engagement recognition model on eight different activities. We compared the accuracy and F1 score of the model before and after applying transfer learning, and observed how the model's performance varied across different activities.

For example, the results show that applying transfer learning has generally improved the accuracy and F1 score of the model for binary classification in most activities. For instance, when the model was pre-trained on the "Touch Me" activity and then fine-tuned on the "Songs" activity subset, the highest improved accuracy and F1 score were achieved, with values of 74.81% and 85.56%, respectively. A similar

trend was observed in the “Dances” and “Storytelling” activities. On the other hand, the accuracy and F1 score for multi-class classification remained mostly the same or slightly decreased while pre-trained on activities such as “Touch Me,” “Imitation,” “Emotions,” and "Hello&Bye." However, in the “Touch Me” and “Imitation” activities, no improvement was observed, and all values decreased after applying transfer learning. Moreover, in the “Touch Me” activity, the lowest accuracy was achieved while the model was pre-trained on the “Follow Me” activity subset.

These findings suggest that the effectiveness of transfer learning depends on various factors, such as the similarity between the pre-training and fine-tuning tasks, the size and diversity of the datasets, and the complexity of the model architecture. Therefore, it is important to carefully select the pre-training and fine-tuning strategies based on the specific task and dataset at hand.

7.2.4 Concluding Remarks

One group of data is used as an initialization for the classification of another group of data through a series of techniques referred to as transfer learning. We have only so far looked into training across all activities in an experiment, followed by fine-tuning the entire network on a particular activity. There are several more transfer learning methods that might be effective.

In conclusion, the results indicate that transfer learning might be a useful method for enhancing the performance of engagement recognition models for binary classification problems, but it may not necessarily lead to better results in multi-class classification tasks. The pre-training activity selected might also have a big influence on the model’s performance. The results can help practitioners and researchers in the field of engagement recognition increase the F1 score and accuracy of their models.

Chapter 8

Conclusions

The structure of the chapter is aimed to provide a clear and concise overview of our research and highlight the main insights and contributions that we have made in the field. In general, this thesis addressed the role of social robots and engagement in long-term CRI. More specifically, we studied the engagement during the interactions between children diagnosed with autism and robots.

8.1 Long-term RAAT

Children with ASD typically experience difficulties in adaptability skills, interpersonal synchrony, self-initiation, and eye contact [14]. They may also resist engaging with human therapists [233], making effective interaction challenging [113, 134]. Recent research indicates that robots are well-received by children with ASD, positively impacting their imitation abilities, eye contact, behavioural reactions, joint attention, and repetitive or stereotyped behaviours [20, 151]. Furthermore, autistic individuals often acquire life skills more effectively when interacting with robots than with humans [17, 64, 151].

Research on autism requires intervention-based experiments that focus on designing and conducting special therapy to assist in the improvements of social and communication skills. However, there is an issue of generalizability, as individuals with autism exhibit varying degrees of symptoms and may respond differently to therapy.

As a result, therapy plans are often customized and tailored to each child's individual needs and preferences.

In this thesis, we conducted a long-term RAT for children with ASD and ADHD, examining an adaptive experience that brought positive changes in children's behaviours through long-term engagement. We investigated whether employing familiar activities for each child could enhance engagement levels across sessions, as hypothesized in *H1* (see Section 1.6). We conducted qualitative and quantitative analyses of interactions of children with autism with the NAO robot during RAAT and found statistically significant differences in all four measures between adaptive and non-adaptive sessions. These findings support the hypothesis that personalized sessions are more effective for children with ASD. As described in Section 4, implementing personalized rehabilitation and assessment techniques for children with ASD and ADHD is crucial. Based on therapist feedback from our study, we suggest designing activities that incorporate a triadic setting, including typically developed children such as the child's sibling [159].

In summary, this research adds to the literature on building an adaptive approach based on ASD children's needs and differences, focusing on the long-term commitment to interventions.

8.2 QAMQOR Dataset

As mentioned in the previous Section 2, engagement is vital in HRI, not just for controlling interaction design and implementation but also for enabling highly sophisticated interactions to respond to users. To indicate the effect of robots within autism therapy and the quality of interaction several projects have measured the engagement of the children [107]. The concept of engagement is interpreted differently [143]. In our work, engagement is defined as one's active and appropriate involvement in the task with the robot and/or the therapist, which includes attention, involvement, interest, immersion, rapport, empathy, and stance [180]. Many scientific works investigated the use of social cues to help and assess children's engagement [148, 213].

However, child behaviour is heterogeneous: the differences in facial expressions, body gestures and tone of voice.

In order to test our *H2* that states, recognition of engagement would be more accurate if the model trains on multimodal data compared to the single type of data, we generated a new QAMQOR dataset (Section 5) to evaluate standard machine learning algorithms for the engagement recognition task. The QAMQOR is a multimodal dataset of behavioural data recorded from 34 children diagnosed with ASD, covering 194 therapy sessions and more than 48 hours of video. We evaluated these multimodal data (facial and body modalities) in conjunction with four different split methods, namely Random, Child Independent, Session Independent, and Activity Independent. The highest percentage of accuracy during testing was attained through examination of OpenPose keypoints, particularly during the Child Independent split, resulting in a binary classification accuracy of 84.17% and a multi-class classification accuracy of 43.21%. After analyzing the outcomes presented in Section 6, we can conclude that body multimodal features demonstrate a higher level of precision when recognizing engagement in both binary and multi-class classification tasks, particularly in the Random split. Conversely, when conducting Child and Session Independent splits, face multimodal features tend to yield better results. The experimental results demonstrate that the accuracy of engagement recognition systems can be improved by selecting an appropriate feature modality depending on the classification problem and split. While full-frame features may perform well in some situations, other modalities may be more effective in addressing specific challenges.

In summary, as for the practical utility of the robots in a clinical environment, the real-time engagement estimation model could allow therapists and caregivers to detect small behavioural changes in children with ASD and keep track of them. In addition, it provides a possibility for the therapists to monitor engagement. Finally, the engagement recognition model will help to label new data and save resources and time during the annotation process.

8.3 Applying Transfer Learning Approach

Data-driven approaches in HRI face challenges due to small and context-specific datasets, especially when dealing with children with autism. Obtaining a sufficient number of manually labelled training samples can be expensive and time-consuming. Nonetheless, there is a need for algorithms that can learn effectively from limited labelled training data by utilizing related labelled data or data with different distributions.

During this PhD research, we investigated the effectiveness of transfer learning in utilizing multimodal features for engagement recognition in children diagnosed with ASD.

The availability of publicly accessible CRI datasets, particularly those featuring interactions between children with autism and robots, is limited, as discussed in Section 2.8. Our research objective was to improve engagement recognition accuracy using the transfer learning approach, by training the model on a comparable dataset and identifying any transferable features between the typical population and the population diagnosed with autism. To achieve this, we utilized the PInSoRo dataset, a publicly available dataset featuring free-play interactions between typically developed children and the NAO robot.

To test our *H3*, we applied transfer learning by pre-training a model on a similar available CRI dataset and then fine-tuning it on the QAMQOR dataset. However, the results (Section 7.1.2) showed that the engagement recognition accuracy decreased after applying transfer learning, which led us to reject our hypothesis. Our results showed that none of the multimodal features was transferable from the typical population to the population diagnosed with ASD. This means that the features that are effective in engagement recognition for typical children are not useful for engagement recognition in children with ASD. These results emphasize the challenges of developing effective engagement recognition models for children with ASD, particularly when dealing with small and context-specific datasets.

It is worth noting that while our experiment did not find transfer learning to be

effective for the engagement recognition of children with ASD, this does not imply that transfer learning is not a useful approach for this population. Further research is needed to explore other potential factors that may affect the effectiveness of transfer learning for engagement recognition in children with ASD.

However, the effectiveness of engagement recognition models may depend on the context of the interaction and the type of activity performed by the robot. This raises questions about the generalizability of engagement recognition models across different robot activities.

We explored the generalizability of engagement recognition models by investigating whether knowledge can be transferred from one activity to another with similar instances. Our experiment aimed to test *H4*, which states that we can transfer knowledge from one activity to another with similar instances (face and body features) using the transfer learning approach for engagement recognition. To achieve this, we used the QAMQOR dataset, consisting of eight different robot activities, including "Songs," "Dances," "Touch Me," "Storytelling," "Imitations," "Emotions," "Follow Me," and "Hello&Bye." Specifically, we pre-trained an engagement recognition model on one activity and evaluated its performance on other activities. By doing so, we aimed to explore whether the knowledge gained from one activity could be applied to another using transfer learning, a popular approach for improving model performance with limited labelled data.

Based on the result, written in Section 7.2.2, transfer learning had a positive impact on the model's accuracy and F1 score in some activities but not in others. For example, when pre-trained on the "Touch Me" activity, the model's performance did not improve when fine-tuned on the "Touch Me" or "Imitation" activity subsets but did improve when fine-tuned on the "Dances" or "Hello&Bye" activity subsets. Similarly, when pre-trained on the "Storytelling" activity, the model's performance improved for binary classification when fine-tuned on all activity subsets, but slightly decreased for multi-class classification.

The findings from this experiment will contribute to a better understanding of the generalizability of engagement recognition models in HRI and provide insights into

developing more effective models that can be applied to various robot activities.

In summary, this thesis contributes to the existing knowledge of engagement recognition in HRI for children with ASD, highlighting the need for tailored approaches and models. Our findings underscore the importance of developing effective engagement recognition models that can improve the quality of RAAT.

8.4 Limitations

While the RAAT shows potential, it is important to consider the limitations and challenges associated with this approach. During the course of PhD research, there are various limitations that need to be taken into account.

- Due to the COVID-19 pandemic, we encountered a limitation in data collection as we were unable to gather the intended sample size of 100 children with autism as originally planned.

- One of the limitations of using robots for therapy is that technical issues and connectivity problems can arise, leading to dropped engagement levels of children during their interaction with the robot.

- Due to time constraints, we were unable to implement a new model for engagement recognition, and instead had to evaluate the QAMQOR dataset using only standard machine learning algorithms.

- The accuracy of the results was limited due to the highly imbalanced and noisy nature of the data, which made it difficult for the machine learning algorithms to accurately classify the data.

- The experiment is limited by the fact that the QAMQOR dataset was utilized in the analysis, but the PInSoRo dataset was used in the trial and was obtained from typically developed children. In addition, there were a significant amount of missing data points in the PInSoRo dataset, which may have impacted the accuracy of the obtained model. In addition, the PInSoRo dataset only included

18 keypoints of the body and lacked key features like keypoints for the left and right hands, which limited the capabilities of the QAMQOR dataset by limiting the features to 18 out of 25 keypoints of the body. The effectiveness of transfer learning depends on the availability of a highly accurate model trained on good data. Therefore, it may not be as effective due to the underlying differences in neural mechanisms for learning and processing information.

- There was still room for further investigation with a larger dataset of children with autism (for example, DREAM), which is necessary to determine the efficacy of transfer learning techniques in this population.

8.5 Future Work

Despite the progress made in engagement recognition using HRI technology for children with autism, several areas for future research exist.

First, while the use of multimodal features has shown promise in improving engagement recognition accuracy, further research is needed to explore the potential of additional modalities, such as physiological and psychological data.

Second, a standardized model for engagement recognition in children with autism would facilitate comparisons between different studies and enable the development of more effective and tailored approaches to therapy.

Third, it is necessary to explore larger and more diverse datasets that consist of data from children with ASD. One possible approach is to use transfer learning with similar datasets, such as the DREAM or MDCA dataset, to develop more robust and generalizable engagement recognition models that can be applied to different settings and scenarios.

Fourth, it is important to experiment with different neural network architectures, including deep learning and recurrent neural networks, to improve the performance of engagement recognition models. This can involve varying the number of hidden layers and nodes in each layer to identify the optimal architecture for the specific task of engagement recognition in children with ASD.

Fifth, while interactive communication technologies show potential for improving the lives of children with autism, more research is required to develop practical and effective approaches to therapy. The future work outlined above will contribute to addressing the challenges faced in engagement recognition using HRI technology and ultimately improve the skills of children with autism.

Overall, this thesis provides a strong foundation for future research in the field of interactive technologies for children with autism, and we anticipate the extension and application of our findings in practice.

Glossary

Adaptive sessions: RAT sessions that contain mostly seen activities by a child and follow each child's familiarity, preferences, and unique characteristics.

Engagement: Individual's level of involvement, interest, attention, and active participation in a particular activity, task, or interaction. A child who is engaged is more likely to benefit from the therapy sessions, as they are actively participating, learning, and building essential social and cognitive skills.

Engagement recognition model: A machine learning model used to recognize and assess the engagement levels of children with autism in the QAMQOR dataset.

Multimodal data: The collection and integration of information from modalities during therapy sessions with children with autism. Our study employs different visual data from different modalities.

Multi-session study: A research that involves interactions with participants over multiple distinct sessions.

QAMQOR dataset: A valuable resource in the context of our research, specifically designed to support the development and evaluation of engagement recognition models for children with autism during therapy sessions.

Transfer learning: A machine learning technique where knowledge gained from one activity/dataset is applied to improve performance on a target activity/dataset to improve engagement recognition models.

Valence: The emotional quality of an individual's response during interactions with a social robot in autism therapy sessions. We use behavioural observations to measure facial expressions, body language, and vocalizations, which provide valuable cues about the child's emotional state.

Bibliography

[1] Treatment and intervention services for autism spectrum disorder. `https://www.cdc.gov/ncbddd/autism/treatment.html`. Accessed: 2022-08-31.

[2] Sevgi Nur Bilgin Aktaş, Pınar Uluer, Buket Coşkun, Elif Toprak, Duygun Erol Barkana, Hatice Köse, Tatjana Zorcec, Ben Robins, and Agnieszka Landowska. Stress detection of children with asd using physiological signals. In *2022 30th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2022.

[3] Alyssa M. Alcorn, Eloise Ainger, Vicky Charisi, Stefania Mantinioti, Sunčica Petrović, Bob R. Schadenberg, Teresa Tavassoli, and Elizabeth Pellicano. Educators' views on using humanoid robots with autistic learners in special education settings in england. *Frontiers in Robotics and AI*, 6:107, 2019.

[4] Anna Altavas. An interdisciplinary team model in diagnosing autism helps brendan find his voice, 2018.

[5] Aida Amirova, Nazerke Rakhymbayeva, Elmira Yadollahi, Anara Sandygulova, and Wafa Johal. 10 years of human-nao interaction research: A scoping review. *Frontiers in Robotics and AI*, 8:744526, 2021.

[6] S. Andrist, B. Mutlu, and A. Tapus. Look like me: Matching robot personality via gaze to increase motivation. *In Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3603–3612, 2015.

[7] Salvatore M Anzalone, Sofiane Boucenna, Serena Ivaldi, and Mohamed Chetouani. Evaluating the engagement with social robots. *International Journal of Social Robotics*, 7(4):465–478, 2015.

[8] Salvatore Maria Anzalone, Elodie Tilmont, Sofiane Boucenna, Jean Xavier, Anne-Lise Jouen, and Nicolas Bodeau. How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3d1 time) environment during a joint attention induction task with a robot. *Research in Autism Spectrum Disorders*, 8:814–826, 2014.

[9] Salvatore Maria Anzalone, Jean Xavier, Sofiane Boucenna, Lucia Billeci, Antonio Narzisi, Filippo Muratori, David Cohen, and Mohamed Chetouani. Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment. *Pattern Recognition Letters*,

118:42 – 50, 2019. Cooperative and Social Robots: Understanding Human Activities and Intentions.

[10] S.M. Anzalone and M. Chetouani. Tracking posture and head movements of impaired people during interactions with robots. *New Trends in Image Analysis and Processing–ICIAP*, Springer Berlin Heidelberg:41–49, 2013.

[11] APA. *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC, 5th ed. edition, 2013.

[12] Syzdykbayev Azamat. The number of children with autism has doubled in kazakhstan. *KazInform*, 13 November 2022.

[13] Kim Baraka, Patrícia Alves-Oliveira, and Tiago Ribeiro. *An Extended Framework for Characterizing Social Robots*, pages 21–64. Springer International Publishing, Cham, 2020.

[14] S. Baron-Cohen and C. Gillberg. Mind blindness: An essay on autism and theory of mind, developmental medicine and child neurology. *Developmental Medicine and Child Neurology*, 37(12):1124, 1995.

[15] Katrin D. Bartl-Pokorny, Małgorzata Pykała, Pinar Uluer, Duygun Erol Barkana, Alice Baird, Hatice Kose, Tatjana Zorcec, Ben Robins, BJörn W. Schuller, and Agnieszka Landowska. Robot-based intervention for children with autism spectrum disorder: A systematic literature review. *IEEE Access*, 9:165433–165450, 2021.

[16] Paul Baxter, Tony Belpaeme, Lola Cañamero, Piero Cosi, Yiannis Demiris, and Valentin Enescu. Long-term human-robot interaction with young users. In *in Proceedings of the ACM/IEEE Human-Robot Interaction conference (HRI-2011) (Robots with Children Workshop*, 2011.

[17] Momotaz Begum, Richard Serna, and Holly Yanco. Are robots ready to deliver autism interventions? a comprehensive review. *International Journal of Social Robotics*, 8, 03 2016.

[18] Esube Bekele, Uttama Lahiri, Amy Swanson, Julie Crittendon, Zachary Warren, and Nilanjan Sarkar. A step towards developing adaptive robot-mediated intervention architecture (aria) for children with autism. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 21, 12 2012.

[19] T. Belpaeme, P.E. Baxter, R. Read, R. Wood, H. Cuayhuitl, B. Kiefer, S. Racioppa, I. Kruijff-Korbayova, G. Athanasopoulos, V. Enescu, R. Looije, M. Neerincx, Y. Demiris, R. Ros-Espinoza, A. Beck, L. Canamero, A. Hiolle, M. Lewis, I. Baroni, M. Nalin, P. Cosi, G. Paci, F. Tesser, G. Sommavilla, and R. Humbert. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1:33–53, 2012.

[20] Iris Berk-Smeekens, Martine Dongen-Boomsma, Manon Korte, Jenny Boer, Iris Oosterling, Nienke Peters-Scheffer, Jan Buitelaar, Emilia Barakova, Tino Lourens, Wouter Staal, and Jeffrey Glennon. Adherence and acceptability of a robot-assisted pivotal response treatment protocol for children with autism spectrum disorder. *Scientific Reports*, 10, 2020.

[21] Jaishankar Bharatharaj, Loulin Huang, Rajesh Elara Mohan, Ahmed Al-Jumaily, and Christian Krägeloh. Robot-assisted therapy for learning and social interaction of children with autism spectrum disorder. *Robotics*, 6(1), 2017.

[22] Anjana Bhat. Is motor impairment in autism spectrum disorder (asd) distinct from developmental coordination disorder (dcd)? a report from the spark study. *Physical therapy*, 100:633–644, 2020.

[23] Erik Billing, Tony Belpaeme, Haibin Cai, Hoang-Long Cao, Anamaria Ciocan, Cristina Costescu, Daniel David, Robert Homewood, Daniel Hernandez Garcia, Pablo Gómez Esteban, Honghai Liu, Vipul Nair, Silviu Matu, Alexandre Mazel, Mihaela Selescu, Emmanuel Senft, Serge Thill, Bram Vanderborght, David Vernon, and Tom Ziemke. The dream dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy. *PLOS ONE*, 15(8):1–15, 08 2020.

[24] S. Bonfiglio, Mohamed Chetouani, Angele Giuliano, Koushik Maharatna, Filippo Muratori, C. Nugent, Cristiano Paggetti, and Giovanni Pioggia. Michelangelo, an european research project exploring new, ict-supported approaches in the assessment and treatment of autistic children. *Neuropsychiatrie de l'Enfance et de l'Adolescence*, 60(5):S33, 07 2012.

[25] V. Bono, A. Narzisi, A.-L. Jouen, E. Tilmont, S. Hommel, W. Jamal, J. Xavier, L. Billeci, K. Maharatna, and M. Wald. Goliah: a gaming platform for home-based intervention in autism–principles and design. *Front. Psychiatry*, 7, 2016.

[26] S. Boucenna, S. Anzalone, E. Tilmont, D. Cohen, and M. Chetouani. Learning of social signatures through imitation game between a robot and a human partner. *Auton. Mental Dev. IEEE Trans.*, 6(3):213–225, 2014.

[27] Sofiane Boucenna, Antonio Narzisi, Elodie Tilmont, Filippo Muratori, Giovanni Pioggia, David Cohen, and Mohamed Chetouani. Interactive technologies for autistic children: A review. *Cognitive Computation*, 6:722–740, 2014.

[28] G. Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[29] Andrea Brivio, Ksenia Rogacheva, Matteo Lucchelli, and Andrea Bonarini. A soft, mobile, autonomous robot to develop skills through play in autistic children. *Paladyn, Journal of Behavioral Robotics*, 12(1):187–198, 2021.

[30] Daniela Bulgarelli and Nicole Bianquin. *3 Conceptual Review of Play*, pages 58–70. De Gruyter Open Poland, 2017.

[31] J-J. Cabibihan, H. Javed, M.Jr. Ang, and S.M. Aljunied. Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. *Int.J.Soc.Robot*, 5(4):593–618, 2013.

[32] H. Cao, G. Van de Perre, J. Kennedy, E. Senft, P. Gómez Esteban, A. De Beir, R. Simut, T. Belpaeme, D. Lefeber, and B. Vanderborght. A personalized and platform-independent behavior control system for social robots in therapy: Development and applications. *IEEE Transactions on Cognitive and Developmental Systems*, 11(3):334–346, Sep. 2019.

[33] Hoang-Long Cao, Pablo G. Esteban, Madeleine Bartlett, Paul Baxter, Tony Belpaeme, Erik Billing, Haibin Cai, Mark Coeckelbergh, Cristina Costescu, Daniel David, Albert De Beir, Daniel Hernandez, James Kennedy, Honghai Liu, Silviu Matu, Alexandre Mazel, Amit Pandey, Kathleen Richardson, Emmanuel Senft, Serge Thill, Greet Van de Perre, Bram Vanderborght, David Vernon, Kutoma Wakanuma, Hui Yu, Xiaolong Zhou, and Tom Ziemke. 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.

[34] Wei Cao, Wenxu Song, Xinge Li, Sixiao Zheng, Ge Zhang, Yanting Wu, Sailing He, Huilin Zhu, and Jiajia Chen. Interaction with social robots: Improving gaze toward face but not necessarily joint attention in children with autism spectrum disorder. *Frontiers in Psychology*, 10:1503, 2019.

[35] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

[36] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7291–7299, 2017.

[37] Ginevra Castellano, André Pereira, Iolanda Leite, Ana Paiva, and Peter W. McOwan. Detecting user engagement with a robot companion using task and social interaction-based features. In *Proceedings of the 2009 International Conference on Multimodal Interfaces*, ICMI-MLMI '09, page 119–126, New York, NY, USA, 2009. Association for Computing Machinery.

[38] Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing*, 10(4):484–497, 2019.

[39] I. Cester, S. Dunne, A. Riera, and G. Ruffini. Enobio: wearable, wireless, 4-channel electrophysiology recording system optimized for dry electrodes. *Proceedings of the Health Conference*, 2008.

[40] C. Chalmers. Robotics and computational thinking in primary school. *International Journal of Child-Computer Interaction*, 17:93–100, 2018.

[41] C. Chang, C. Zhang, L. Chen, and Y. Liu. An ensemble model using face and body tracking for engagement detection. *In Proceedings of the International Conference on Multimodal Interaction*, ACM:616–622, 2018.

[42] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'16, pages 785–794, New York, NY, USA, 2016. ACM.

[43] Caitlyn Clabaugh, David Becerra, Eric Deng, Gisele Ragusa, and Maja Matari-undefined. Month-long, in-home case study of a socially assistive robot for children with autism spectrum disorder. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18, page 87–88, New York, NY, USA, 2018. Association for Computing Machinery.

[44] Caitlyn Clabaugh, Kartik Mahajan, Shomik Jain, Roxanna Pakkar, David Becerra, Zhonghao Shi, Eric Deng, Rhianna Lee, Gisele Ragusa, and Maja Matarić. Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. *Frontiers in Robotics and AI*, 6:110, 2019.

[45] Alexandre Coninx, Paul Baxter, Elettra Oleari, Sara Bellini, Bert Bierman, Olivier Blanson Henkemans, Lola Cañamero, Piero Cosi, Valentin Enescu, Raquel Ros Espinoza, Antoine Hiolle, Rémi Humbert, Bernd Kiefer, Ivana Kruijff-Korbayová, Rose-marijn Looije, Marco Mosconi, Mark Neerincx, Giulio Paci, Georgios Patsis, Clara Pozzi, Francesca Sacchitelli, Hichem Sahli, Alberto Sanna, Giacomo Sommavilla, Fabio Tesser, Yiannis Demiris, and Tony Belpaeme. Towards long-term social child-robot interaction: Using multi-activity switching to engage young users. *Journal of Human-Robot Interaction*, 5(1):32–67, 2016.

[46] Lee J. Corrigan, Christopher Peters, Dennis Küster, and Ginevra Castellano. *Engagement Perception and Generation for Social Robots and Virtual Agents*, pages 29–51. Springer International Publishing, Cham, 2016.

[47] L.J. Corrigan, Christopher Peters, Ginevra Castellano, Fotis Papadopoulos, Aidan Jones, Shweta Bhargava, Srini Janarthanam, Helen Hastie, Amol Deshmukh, and Ruth Aylett. Social-task engagement: Striking a balance between the robot and the task. 01 2013.

[48] Andreia P. Costa, Louise Charpiot, Francisco Rodríguez Lera, Pouyan Ziafati, Aida Nazarikhorram, Leendert Van Der Torre, and Georges Steffgen. More attention and less repetitive and stereotyped behaviors using a robot with children with autism. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 534–539, 2018.

[49] S. Costa, H. Lehmann, K. Dautenhahn, B. Robins, and F. Soares. Using a humanoid robot to elicit body awareness and appropriate physical interaction in children with autism. *International journal of social robotics*, 7(2):265–278, 2015.

[50] S. Costa, H. Lehmann, B. Robins, K. Dautenhahn, and F. Soares. "where is your nose?" - developing body awareness skills among children with autism using a humanoid robot. *The Sixth International conference on Advances in Computer-Human Interactions*, Nice, France:117–122, 24 February - 1 March 2013.

[51] Sandra Costa, Filomena Soares, Ana Paula Pereira, Cristina Santos, and Antoine Hiolle. Building a game scenario to encourage children with autism to recognize and label emotions using a humanoid robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 820–825, 2014.

[52] Despoina Damianidou, Ami Eidels, and Michael Arthur-Kelly. The use of robots in social communications and interactions for individuals with asd: A systematic review. *Advances in Neurodevelopmental Disorders*, 4, 12 2020.

[53] K. Dautenhahn, C.L. Nehaniv, M.L. Walters, B. Robins, H. Kose-Bagci, N.A. Mirza, and Blow M. Kaspar-a minimally expressive humanoid robot for human-robot interaction research. *Journal of Applied Bionics and Biomechanics*, 6:369–397, 2009.

[54] Kerstin Dautenhahn. I could be you: the phenomenological dimension of social understanding. *Cybernetics and Systems*, 28(5):417–453, 1997.

[55] Kerstin Dautenhahn. Roles and functions of robots in human society: implications from research in autism therapy. *Robotica*, 21(4):443–452, 2003.

[56] Kerstin Dautenhahn and Iain Werry. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12:1–35, 06 2004.

[57] Kerstin Dautenhahn and Iain P. Werry. Towards interactive robots in autism therapy: background, motivation and challenges. *Pragmatics & Cognition*, 12:1–35, 2004.

[58] Daniel O David, Cristina A Costescu, Silviu Matu, Aurora Szentagotai, and Anca Dobrean. Effects of a robot-enhanced intervention for children with asd on teaching turn-taking skills. *Journal of Educational Computing Research*, page 0735633119830344, 2019.

[59] M. Davis, K. Dautenhahn, C.L. Nehaniv, and S.D. Powell. Guidelines for researchers and practitioners designing software and software trials for children with autism. *Journal of Assistive Technologies*, 4(1):38–48, 2009.

[60] M. Davis, N. Otero, K. Dautenhahn, C. Nehaniv, and S. Powell. Creating a software to promote understanding about narrative in children with autism. *Proc. 6th IEEE International Conference on Development and Learning*, Imperial College, London:64–69, 11-13 July 2007.

[61] Manon de Korte, Iris Berk-Smeekens, Martine Dongen-Boomsma, Iris Oosterling, Jenny Boer, Emilia Barakova, Tino Lourens, Jan Buitelaar, Jeffrey Glennon, and Wouter Staal. Self-initiations in young children with autism during pivotal response treatment with and without robot assistance. *Autism*, 24, 07 2020.

[62] Marco Del Coco, Marco Leo, Pierluigi Carcagnì, Francesca Famà, Letteria Spadaro, Liliana Ruta, Giovanni Pioggia, and Cosimo Distante. Study of mechanisms of social interaction stimulation in autism spectrum disorder by assisted humanoid robot. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):993–1004, 2018.

[63] P. Dickerson, B. Robins, and K. Dautenhahn. Where the action is: A conversation analytic perspective on interaction between a humanoid robot, a co-present adult and a child with an asd. *Interaction Studies*, 14(2):297–316, 2013.

[64] Joshua J Diehl, Lauren M Schmitt, Michael Villano, and Charles R Crowell. The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in autism spectrum disorders*, 6(1):249–262, 2012.

[65] Sidney S D'Mello, Patrick Chipman, and Art Graesser. Posture as a predictor of learner's affective engagement. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29:571–576, 2007.

[66] Jane Doussard-Roosevelt, Claudia Joe, Olga Bazhenova, and Stephen Porges. Mother-child interaction in autistic and nonautistic children: Characteristics of maternal approach behaviors and child social responses. *Development and psychopathology*, 15:277–95, 2003.

[67] Karl Drejing, Serge Thill, and Paul Hemeren. Engagement: A traceable motivational concept in human-robot interaction. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 956–961, 2015.

[68] Audrey Duquette, François Michaud, and Henri Mercier. Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots*, 24:147–157, 2008.

[69] Paul Ekman and Wallace V. Friesen. Facial action coding system: a technique for the measurement of facial movement. 1978.

[70] David Feil-Seifer and Maja J. Mataric. B3ia: A control architecture for autonomous robot-assisted behavior intervention for children with autism spectrum disorders. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 328–333, 2008.

[71] David Feil-Seifer and Maja J. Matarić. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In Oussama Khatib, Vijay Kumar, and George J. Pappas, editors, *Experimental Robotics*, pages 201–210, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[72] David Feil-Seifer and Maja J. Matarić. Socially assistive robotics. *IEEE Robotics Automation Magazine*, 18(1):24–31, 2011.

[73] Yongli Feng, Jia Qinqxuan, and Wei Wei. A control architecture of robot-assisted intervention for children with autism spectrum disorders. *Journal of Robotics*, 1:1–13, 2018.

[74] Ester Ferrari, Ben Robins, and Kerstin Dautenhahn. Therapeutic and educational objectives in robot assisted play for children with autism. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 108–114, 2009.

[75] M. Filippi and F. Agosta. *Handbook of Clinical Neurology*, volume 136. 2016.

[76] Marino Flavia, P. Chilà, Stefania Sfrazzetto, Cristina Carrozza, Ilaria Crimi, Chiara Failla, Mario Busà, Giuseppe Bernava, Gennaro Tartarisco, David Vagni, Liliana Ruta, and Giovanni Pioggia. Outcomes of a robot-assisted social-emotional understanding intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 50:1973–1987, 2020.

[77] D. François, D. Polani, and Dautenhahn K. Towards socially adaptive robots: A novel method for real time recognition of human-robot interaction styles. *Proc. IEEE-RAS International Conference on Humanoid Robots*, Daejeon, South Korea, 2008.

[78] Dorothée François, Stuart Powell, and Kerstin Dautenhahn. A long-term study of children with autism playing with a robotic pet: Taking inspirations from non-directive play therapy to encourage children's proactivity and initiative-taking. *Interaction Studies*, 10:324–373, 2009.

[79] Jennifer A Fredricks, Phyllis C Blumenfeld, and Alison H Paris. School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1):59–109, 2004.

[80] Jennifer A. Fredricks and Wendy McColskey. *The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments*, pages 763–782. Springer US, Boston, MA, 2012.

[81] Kristin Fuglerud and Ivar Solheim. The use of social robots for supporting language training of children. 10 2018.

[82] Rossella Gambetti and Guendalina Graffigna. The concept of engagement. *International Journal of Market Research*, 52:801–826, 01 2010.

[83] Michael Gazzaniga, Richard Ivry, and George Mangun. *Cognitive Neuroscience: The Biology of the Mind*. 01 2013.

[84] Stefano Ghidoni, Salvatore M. Anzalone, Matteo Munaro, Stefano Michieletto, and Emanuele Menegatti. A distributed perception infrastructure for robot assisted living. *Robotics and Autonomous Systems*, 62(9):1316–1328, 2014. Intelligent Autonomous Systems.

[85] Irini Giannopulu, Valérie Montreynaud, and Tomio Watanabe. Neurotypical and autistic children aged 6 to 7 years in a speaker-listener situation with a human or a minimalist interactor robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 942–948, 2014.

[86] Nadine Glas and Catherine Pelachaud. Definitions of engagement in human-agent interaction. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 944–949, 2015.

[87] S. Golestan, P. Soleiman, and H. Moradi. A comprehensive review of technologies used for screening, assessment, and rehabilitatioin of autism spectrum disorder. 2017.

[88] J. Greczek, E. Kaszubski, A. Atrash, and M. Matarić. Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 561–566, Aug 2014.

[89] Ivan Grishchenko and Valentin Bazarevsky. Mediapipe holistic — simultaneous face, hand and pose prediction, on device.

[90] O.A.B. Henkemans, B.P. Bierman, J. Janssen, M.A. Neerincx, R. Looije, van der Bosch H., and van der Giessen J.A. Using a robot to personalise health education for children with diabetes type 1: A pilot study. *Journal of Patient education and counseling*, 92:174–181, 2013.

[91] Claire AGJ Huijnen, Monique AS Lexis, and Luc P de Witte. Matching robot kaspar to autism spectrum disorder (asd) therapy and educational goals. *International Journal of Social Robotics*, 8(4):445–455, 2016.

[92] Bibi Huskens, Annemiek Palmen, Marije Werff, Tino Lourens, and Emilia Barakova. Improving collaborative play between children with autism spectrum disorders and their siblings: The effectiveness of a robot-mediated intervention based on lego® therapy. *Journal of autism and developmental disorders*, 45, 2014.

[93] Susan L. Hyman, Susan E. Levy, Scott M. Myers, SECTION ON DEVELOPMENTAL COUNCIL ON CHILDREN WITH DISABILITIES, BEHAVIORAL PEDIATRICS, Dennis Z. Kuo, Susan Apkon, Lynn F. Davidson, Kathryn A.

Ellerbeck, Jessica E.A. Foster, Garey H. Noritz, Mary O'Connor Leppert, Barbara S. Saunders, Christopher Stille, Larry Yin, Carol C. Weitzman, Jr Childers, David Omer, Jack M. Levine, Ada Myriam Peralta-Carcelen, Jennifer K. Poon, Peter J. Smith, Nathan Jon Blum, John Ichiro Takayama, Rebecca Baum, Robert G. Voigt, and Carolyn Bridgemohan. Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. *Pediatrics*, 145(1), 01 2020. e20193447.

[94] I. Iacono, H. Lehmann, P. Marti, B. Robins, and K. Dautenhahn. Robots as social mediators for children with autism - a preliminary analysis comparing two different robotic platforms. *first Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*, Frankfurt am Main, Germany:1–6, 24-27 August 2011.

[95] Koji Inoue, Divesh Lala, Katsuya Takanashi, and Tatsuya Kawahara. Engagement recognition by a latent character model based on multimodal listener behaviors in spoken dialogue. *APSIPA Transactions on Signal and Information Processing*, 7:e9, 2018.

[96] L.I. Ismail, T. Verhoeven, J. Dambre, and F. Wyffels. Leveraging robotics research for children with autism: A review. *International Journal of Social Robotics*, 2018.

[97] 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.

[98] J.B. Janssen, C.C. van der Wal, M.A. Neerincx, and R. Looije. Motivating children to learn arithmetic with an adaptive robot game. *Springer*, 2011.

[99] Robert M. Joseph, Helen Tager-Flusberg, and Catherine Lord. Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of child psychology and psychiatry, and allied disciplines*, 43 6:807–21, 2002.

[100] Debra Kamps, Rose Mason, and Linda Heitzman-Powell. *Peer Mediation Interventions to Improve Social and Communication Skills for Children and Youth with Autism Spectrum Disorders*, pages 257–283. 2017.

[101] Takayuki Kanda, Rumi Sato, Naoki Saiwaki, and Hiroshi Ishiguro. A two-month field trial in an elementary school for long-term human–robot interaction. *Robotics, IEEE Transactions on*, 23:962 – 971, 11 2007.

[102] Ashish Kapoor and Rosalind W. Picard. Multimodal affect recognition in learning environments. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, MULTIMEDIA '05, page 677–682, New York, NY, USA, 2005. Association for Computing Machinery.

[103] Deb Keen. Engagement of children with autism in learning. *Australasian Journal of Special Education*, 33(2):130–140, March 2012.

[104] Saskia M. Kelders, Llewellyn Ellardus van Zyl, and Geke D. S. Ludden. The concept and components of engagement in different domains applied to ehealth: A systematic scoping review. *Frontiers in Psychology*, 11, 2020.

[105] J. Kennedy, P. Baxter, and T. Belpaeme. The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning. *In Proc. 10th ACM/IEEE International Conference on Human-Robot Interaction*, pages 67–74, 2015.

[106] Elizabeth Kim, Christopher Daniell, Corinne Makar, Julia Elia, Brian Scassellati, and Frederick Shic. Potential clinical impact of positive affect in robot interactions for autism intervention. pages 8–13, 09 2015.

[107] Elizabeth Kim, Rhea Paul, Frederick Shic, and Brian Scassellati. Bridging the research gap: Making hri useful to individuals with autism. *Journal of Human-Robot Interaction*, 1, 08 2012.

[108] Elizabeth S Kim, Lauren D Berkovits, Emily P Bernier, Dan Leyzberg, Frederick Shic, Rhea Paul, and Brian Scassellati. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of autism and developmental disorders*, 43(5):1038–1049, 2013.

[109] Min-Gyu Kim, Iris Oosterling, Tino Lourens, Wouter Staal, Jan Buitelaar, Jeffrey Glennon, Iris Smeekens, and Emilia Barakova. Designing robot-assisted pivotal response training in game activity for children with autism. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1101–1106, 2014.

[110] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[111] Yuriko Kishida and Coral Kemp. The engagement and interaction of children with autism spectrum disorder in segregated and inclusive early childhood center-based settings. *Topics in Early Childhood Special Education*, 29(2):105–118, 2009.

[112] L. K. Koegel, K. M. Bryan, P. L. Su, M. Vaidya, and S. Camarata. Definitions of nonverbal and minimally verbal in research for autism: A systematic review of the literature. *Journal of Autism and Developmental disorders*, 50:2957–2972, 2020.

[113] Lynn Koegel, Cynthia Carter, and Robert Koegel. Teaching children with autism self-initiations as a pivotal response. *Topics in Language Disorders - TOP LANG DISORD*, 23:134–145, 04 2003.

[114] K Kompatsiari, F Ciardo, D De Tommaso, and A Wykowska. Measuring engagement elicited by eye contact in human-robot interaction. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6979–6985, 2019.

[115] Manon WP De Korte, Iris van den Berk-Smeekens, Martine van Dongen-Boomsma, Iris J Oosterling, Jenny C Den Boer, Emilia I Barakova, Tino Lourens, Jan K Buitelaar, Jeffrey C Glennon, and Wouter G Staal. Self-initiations in young children with autism during pivotal response treatment with and without robot assistance. *Autism*, 24(8):2117–2128, 2020. PMID: 32730096.

[116] Athanasia Kouroupa, Keith R. Laws, Karen Irvine, Silvana E. Mengoni, Alister Baird, and Shivani Sharma. The use of social robots with children and young people on the autism spectrum: A systematic review and meta-analysis. *PLOS ONE*, 17(6):1–25, 06 2022.

[117] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive robots for communication-care: a case-study in autism therapy. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication*, pages 341–346, Aug 2005.

[118] Hideki Kozima, Marek Piotr Michalowski, and Cocoro Nakagawa. Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics*, 1(1), January 2009.

[119] Hidekiand Kozima, Cocoro Nakagawa, and Yuriko Yasuda. Children–robot interaction: a pilot study in autism therapy. In C. [von Hofsten] and K. Rosander, editors, *From Action to Cognition*, volume 164 of *Progress in Brain Research*, pages 385 – 400. Elsevier, 2007.

[120] Hirokazu Kumazaki, Taro Muramatsu, Yuichiro Yoshikawa, Yoshio Matsumoto, Masutomo Miyao, Hiroshi Ishiguro, Masaru Mimura, Yoshio Minabe, and Mitsuru Kikuchi. How the realism of robot is needed for individuals with autism spectrum disorders in an interview setting. *Frontiers in Psychiatry*, 10, 2019.

[121] Hirokazu Kumazaki, Yuichiro Yoshikawa, Yuko Yoshimura, Takashi Ikeda, Chiaki Hasegawa, Daisuke Saito, Sara Tomiyama, Kyung-Min An, Jiro Shimaya, Hiroshi Ishiguro, Yoshio Matsumoto, Yoshio Minabe, and Mitsuru Kikuchi. The impact of robotic intervention on joint attention in children with autism spectrum disorders. *Molecular Autism*, 9, 12 2018.

[122] Agnieszka Landowska and Ben Robins. *Robot Eye Perspective in Perceiving Facial Expressions in Interaction with Children with Autism*, pages 1287–1297. 03 2020.

[123] Justin B. Leaf, Joseph H. Cihon, Julia L. Ferguson, and Sara M. Weinkauf. *An Introduction to Applied Behavior Analysis*, pages 25–42. Springer International Publishing, Cham, 2017.

[124] Susan Leekam, Carmen Nieto, Sarah Libby, Lorna Wing, and Judith Gould. Describing the sensory abnormalities of children and adults with autism. *Journal of autism and developmental disorders*, 37:894–910, 2007.

[125] P. Leijdekkers, V. Gay, and F. Wong. Capturemyemotion: A mobile app to improve emotion learning for autistic children using sensors. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 381–384, June 2013.

[126] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, 5, 2013.

[127] Severin Lemaignan, Fernando Garcia, Alexis Jacq, and Pierre Dillenbourg. From real-time attention assessment to "with-me-ness" in human-robot interaction. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 157–164, 2016.

[128] Séverin Lemaignan, Charlotte E. R. Edmunds, Emmanuel Senft, and Tony Belpaeme. The pinsoro dataset: Supporting the data-driven study of child-child and child-robot social dynamics. *PLOS ONE*, 13(10):1–19, 10 2018.

[129] Kent R. Logan, Roger Bakeman, and Elizabeth B. Keefe. Effects of instructional variables on engaged behavior of students with disabilities in general education classrooms. *Exceptional Children*, 63(4):481–497, 1997.

[130] Catharine Lord, Michaele Rutter, P DiLavore, S Risi, K Gotham, S Bishop, et al. Autism diagnostic observation schedule–2nd edition (ados-2). *Los Angeles, CA: Western Psychological Corporation*, 284, 2012.

[131] Catherine Lord, Mayada Elsabbagh, Gillian Baird, and Jeremy Veenstra-Vanderweele. Autism spectrum disorder. *The Lancet*, 392:508–520, Aug 11 2018.

[132] Kelly A. Maenner, Matthew J.and Shaw, Amanda V. Bakian, Deborah A. Bilder, Maureen S. Durkin, Amy Esler, Sarah M. Furnier, Libby Hallas, Jennifer Hall-Lande, Hudson Allison, Michelle M. Hughes, Mary Patrick, Karen Pierce, Jenny N. Poynter, Angelica Salinas, Josephine Shenouda, Alison Vehorn, Zachary Warren, John N. Constantino, Monica DiRienzo, Robert T. Fitzgerald, Andrea Grzybowski, Margaret H. Spivey, Sydney Pettygrove, Walter Zahorodny, Akilah Ali, Jennifer G. Andrews, Thaer Baroud, Johanna Gutierrez, Amy Hewitt, Li-Ching Lee, Maya Lopez, Kristen Clancy Mancilla, Dedria McArthur, Yvette D. Schwenk, Anita Washington, Susan Williams, and Mary E. Cogswell. Prevalence and characteristics of autism spectrum disorder among

children aged 8 years — autism and developmental disabilities monitoring network. *MMWR Surveill Summ 2021*, 11(70):1–16, 2018.

[133] Kerry Magro. Autism is one word attempting to describe millions of different stories, Feb 12 2020. Accessed: 2022-08-31.

[134] Kerry L. Marsh, Robert W. Isenhower, Michael J. Richardson, Molly Helt, Alyssa D. Verbalis, R. C. Schmidt, and Deborah Fein. Autism and social disconnection in interpersonal rocking. *Front Integr Neurosci*, 7(4), 2013.

[135] A. Mazel and S. Matu. Dream lite: Simplifying robot assisted therapy for asd. 2019.

[136] R. A. McWilliam and Jr. Donald B. Bailey. Effects of classroom social structure and disability on engagement. *Topics in Early Childhood Special Education*, 15(2):123–147, 1995.

[137] Francisco S. Melo, Alberto Sardinha, David Belo, Marta Couto, Miguel Faria, Anabela Farias, Hugo Gambôa, Cátia Jesus, Mithun Kinarullathil, Pedro Lima, Luís Luz, André Mateus, Isabel Melo, Plinio Moreno, Daniel Osório, Ana Paiva, Jhielson Pimentel, João Rodrigues, Pedro Sequeira, Rubén Solera-Ureña, Miguel Vasco, Manuela Veloso, and Rodrigo Ventura. Project inside: towards autonomous semi-unstructured human–robot social interaction in autism therapy. *Artificial Intelligence in Medicine*, 96:198 – 216, 2019.

[138] François Michaud and Serge Caron. Roball, the rolling robot. *Auton. Robots*, 12:211–222, 03 2002.

[139] Manuel Milling, Alice Baird, Katrin D. Bartl-Pokorny, Shuo Liu, Alyssa M. Alcorn, Jie Shen, Teresa Tavassoli, Eloise Ainger, Elizabeth Pellicano, Maja Pantic, Nicholas Cummins, and Björn W. Schuller. Evaluating the impact of voice activity detection on speech emotion recognition for autistic children. *Frontiers in Computer Science*, 4, 2022.

[140] AS Mohamed, N Marbukhari, and H Habibah. A deep learning approach in robot-assisted behavioral therapy for autistic children. *International Journal of Advanced Trends in Computer Science and Engineering*, 8(1.6):437–443, 2019.

[141] Marco Nalin, Paul Baxter, Tony Belpaeme, Lola Cañamero, Piero Cosi, Yiannis Demiris, Antoine Hiolle, Ivana Kruijff-Korbayova, Rosemarijn Looije, mark neerikncx, Hichem Sahli, Giacomo Sommavilla, Fabio Tesser, and Rachel Wood. Long-term human-robot interaction with young users. 2011.

[142] Heather L. O'Brien and Elaine G. Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6):938–955, 2008.

[143] Catharine Oertel, Ginevra Castellano, Mohamed Chetouani, Jauwairia Nasir, Mohammad Obaid, Catherine Pelachaud, and Christopher Peters. Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 7:92, 2020.

[144] Marieke Otterdijk, Manon de Korte, I. Smeekens, Jorien Hendrix, Martine Dongen-Boomsma, Jan Buitelaar, Tino Lourens, Jeffrey Glennon, Wouter Staal, and Emilia Barakova. The effects of long-term child–robot interaction on the attention and the engagement of children with autism. *Robotics*, 9:79, 2020.

[145] R. Pakkar, C. Clabaugh, R. Lee, E. Deng, and M. J. Matarićć. Designing a socially assistive robot for long-term in-home use for children with autism spectrum disorders. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7, Oct 2019.

[146] G. Palestra, B. Carolis, and F. Esposito. Artificial intelligence for robot-assisted treatment of autism. *Workshop on Artificial Intelligence with Application in Health, Bari, Italy.*, November 14 2017.

[147] Susan E. Palsbo and Pamela Hood-Szivek. Effect of Robotic-Assisted Three-Dimensional Repetitive Motion to Improve Hand Motor Function and Control in Children With Handwriting Deficits: A Nonrandomized Phase 2 Device Trial. *The American Journal of Occupational Therapy*, 66(6):682–690, 11 2012.

[148] Fotios Papadopoulos, Lee J. Corrigan, Aidan Jones, and Ginevra Castellano. Learner modelling and automatic engagement recognition with robotic tutors. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 740–744, 2013.

[149] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[150] 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.

[151] P. Pennisi, A. Tonacci, G. Tartarisco, L. Billeci, L. Ruta, S. Gangemi, and Pioggia G. Autism and social robotics: A systematic review. *Autism Research*, 9:165–183, 2016.

[152] Olga Perski, Ann Blandford, Robert West, and Susan Michie. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2):254–267, 12 2016.

[153] Christopher Peters, Stylianos Asteriadis, and Kostas Karpouzis. Investigating shared attention with a virtual agent using a gaze-based interface. *Journal on Multimodal User Interfaces*, 3:119–130, 04 2012.

[154] G. Pioggia, M.L. Sica, M. Ferro, R. Igliozzi, F. Muratori, A. Ahluwalia, and D. De Rossi. Human-robot interaction in autism: Face, an android-based social therapy. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 605–612, 2007.

[155] Nicole Poltorak and Alin Drimus. Human-robot interaction assessment using dynamic engagement profiles. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pages 649–654, 2017.

[156] A. Potamianos, C. Tzafestas, E. Iosif, F. Kirstein, P. Maragos, K. Dauthenhahn, J. Gustafson, J-E. Ostergaard, S. Kopp, P. Wik, Pietquin O., and S. Al Moubayed. Babyrobot – next generation social robots: Enhancing communication and collaboration development of td and asd children by developing and commercially exploiting the next generation of human-robot interaction technologies. *2nd Workshop on Evaluating Child-Robot Interaction (CRI) at Human-Robot Interaction*, 7 March 2016.

[157] Michel Puyon and Irini Giannopulu. Emergent emotional and verbal strategies in autism are based on multimodal interactions with toy robots in free spontaneous game play. In *22nd IEEE International Symposium on Robot and Human Interactive Communication: "Living Together, Enjoying Together, and Working Together with Robots!", IEEE RO-MAN 2013*, pages 593–597, United States, 2013. IEEE Computer Society. Co-Chair. Invitation: Professor Tomio Watanabe; 22nd IEEE International Symposium on Robot and Human Interactive Communication: "Living Together, Enjoying Together, and Working Together with Robots!", IEEE RO-MAN 2013 ; Conference date: 26-08-2013 Through 29-08-2013.

[158] Shyam Sundar Rajagopalan, O.V. Ramana Murthy, Roland Goecke, and Agata Rozga. Play with me — measuring a child's engagement in a social interaction. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 1, pages 1–8, 2015.

[159] Nazerke Rakhymbayeva, Aida Amirova, and Anara Sandygulova. A long-term engagement with a social robot for autism therapy. *Frontiers in robotics and AI*, page 669972, 2021.

[160] Anastasia Raptopoulou, Antonios Komnidis, Panagiotis D. Bamidis, and Alexandros Astaras. Human–robot interaction for social skill development in

children with asd: A literature review. *Healthcare Technology Letters*, 8(4):90–96, 2021.

[161] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.

[162] C. Rich, B. Ponsler, A. Holroyd, and C. L. Sidner. Recognizing engagement in human-robot interaction. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 375–382, March 2010.

[163] Daniel Ricks and Mark Colton. Trends and considerations in robot-assisted autism therapy. In *IEEE International Conference on Robotics and Automation*, pages 4354 – 4359, 06 2010.

[164] Giuseppe Riva. Babyrobot: Child–robot communication and collaboration. *Cyberpsychology, Behavior, and Social Networking*, 20(10):660–660, 2017. PMID: 29039702.

[165] Ann S. Roberts, Stephen Shore, Yumiko Mori, Emily Nazzaro, and John Maina. Playing with a robot: Enhancing social communication interaction. In *RESEARCH POSTER PRESENTATION DESIGN*, 2012.

[166] Caroline Robertson and Simon Baron-Cohen. Sensory perception in autism. *Nature reviews. Neuroscience*, 18:671–684, 09 2017.

[167] B. Robins. *A humanoid robot as assistive technology for encouraging social interaction skills in children with Autism*. PhD thesis, University of Hertfordshire, 2005.

[168] B. Robins, K. Dautenhahn, and P. Dickerson. From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. *In Proc. 2nd IEEE International Conferences on Advances in Computer-Human Interactions*, 15:205–211, 2009.

[169] B. Robins, K. Dautenhahn, and P. Dickerson. Embodiment and cognitive learning – can a humanoid robot help children with autism to learn about tactile social behaviour? *International Conference on Social Robotics*, Chengdu, China:66–75, October 2012.

[170] B. Robins, K. Dautenhahn, Te Boekhorst R., and Billard A. Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4:105–120, 2005.

[171] B. Robins, K. Dautenhahn, R. te Boekhorst, and A. Billard. Effects of repeated exposure to a humanoid robot on children with autism. pages 225–236, 2004.

[172] B. Robins, E. Ferrari, and K. Dautenhahn. Developing scenarios for robot assisted play. *In Robot and Human Interactive Communication, RO-MAN The 17th Institute of Electrical and Electronics Engineers (IEEE) International Symposium on*, page 180–186, August 2008.

[173] Ben Robins, Kerstin Dautenhahn, Rene Boekhorst, and Aude Billard. Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4:105–120, 01 2005.

[174] Ben Robins, Nuno Otero, Ester Ferrari, and Kerstin Dautenhahn. Eliciting requirements for a robotic toy for children with autism - results from user panels. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 101 – 106, 09 2007.

[175] José L. Rodrigues, Nuno Gonçalves, Sandra Costa, and Filomena Soares. Stereotyped movement recognition in children with asd. *Sensors and Actuators A: Physical*, 202:162–169, 2013. Selected Papers from the 26th European Conference on Solid-State Transducers Kraków, Poland, 9-12 September 2012.

[176] 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*, 3, June 2018.

[177] O. Rudovic, Y. Utsumi, J. Lee, J. Hernandez, E. Castello Ferrer, B. Schuller, and R. Picard. Culturenet: A deep learning approach for engagement intensity estimation from face images of children with autism. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018.

[178] 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.

[179] Ognjen (Oggi) Rudovic, Jaeryoung Lee, Lea Mascarell-Maricic, Björn W. Schuller, and Rosalind W. Picard. Measuring engagement in robot-assisted autism therapy: A cross-cultural study. *Front. Robot. AI*, 4(36), July 2017.

[180] Hanan Salam, Oya Celiktutan, Isabelle Hupont, Hatice Gunes, and Mohamed Chetouani. Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access*, 14:1–1, 2016.

[181] Mohd.A. Saleh, Fazah Hanapiah, and Habibah Hashim. Robot applications for autism: a comprehensive review. *Disability and Rehabilitation: Assistive Technology*, 16:1–23, 07 2020.

[182] Zohreh Salimi, Ensiyeh Jenabi, and Saeid Bashirian. Are social robots ready yet to be used in care and therapy of autism spectrum disorder: A systematic review of randomized controlled trials. *Neuroscience Biobehavioral Reviews*, 129:1–16, 2021.

[183] Tamie Salter, Neil Davey, and François Michaud. Designing developing queball, a robotic device for autism therapy. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 574–579, 2014.

[184] Anara Sandygulova, Zhanel Zhexenova, Bolat Tleubayev, Aidana Nurakhmetova, Dana Zhumabekova, Ilyas Assylgali, Yerzhan Rzagaliyev, and Aliya Zhakenova. Interaction design and methodology of robot-assisted therapy for children with severe asd and adhd. *Paladyn, Journal of Behavioral Robotics*, 10(1):330–345, 2019.

[185] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P.W. McOwan, and A. Paiva. Automatic analysis of affective postures and body motion to detect engagement with a game companion. *In Proceedings of the 6th international conference on Human-robot interaction,*, ACM:305–312, 2011.

[186] Sunagül Sani-Bozkurt and Gulden Bozkus. Social robots for joint attention development in autism spectrum disorder: A systematic review. *International Journal of Disability, Development and Education*, pages 1–19, 04 2021.

[187] Brian Scassellati, Henny Admoni, and Maja Matarić. Robots for use in autism research. *Annual review of biomedical engineering*, 14:275–94, 05 2012.

[188] Brian Scassellati, Laura Boccanfuso, Chien-Ming Huang, Marilena Mademtzi, Meiying Qin, Nicole Salomons, Pamela Ventola, and Frederick Shic. Improving social skills in children with asd using a long-term, in-home social robot. *Science Robotics*, 3(21), 2018.

[189] Hannah Schertz and Samuel Odom. Promoting joint attention in toddlers with autism: A parent-mediated developmental model. *Journal of autism and developmental disorders*, 37:1562–75, 2007.

[190] Julia Schwarz, Charles Claudius Marais, Tommer Leyvand, Scott E. Hudson, and Jennifer Mankoff. Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 3443–3452, New York, NY, USA, 2014. Association for Computing Machinery.

[191] Emmanuel Senft, Paul Baxter, James Kennedy, and Tony Belpaeme. Sparc: Supervised progressively autonomous robot competencies. In Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi, editors, *Social Robotics*, pages 603–612, Cham, 2015. Springer International Publishing.

[192] Sofia Serholt, Wolmet Barendregt, Asimina Vasalou, Patrícia Alves-Oliveira, Aidan Jones, Sofia Petisca, and Ana Paiva. The case of classroom robots: teachers' deliberations on the ethical tensions. *AI SOCIETY*, 32:613–631, 11 2017.

[193] Jay Sevin, Robert Rieske, and Johnny Matson. A review of behavioral strategies and support considerations for assisting persons with difficulties transitioning from activity to activity. *Review Journal of Autism and Developmental Disorders*, 2:329–342, 2015.

[194] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+d: A large scale dataset for 3d human activity analysis, 2016.

[195] S. Shamsuddin, H. Yussof, F.A. Hanapiah, and L. Ismail. Stereotyped behaviour of autistic children with lower iq level in hri with a humanoid robot. *In Advanced Robotics and its Social Impacts*, IEEE Workshop:175–180, November 2013.

[196] S. Shamsuddin, H. Yussof, L. Ismail, F.A. Hanapiah, S. Mohamed, H.A. Piah, and N.I. Zahari. Initial response of autistic children in human-robot interaction therapy with humanoid robot nao. *In Proc. 8th IEEE International Colloquium on Signal Processing and its Applications*, 16:188–193, 2012.

[197] T Shibata. An overview of human interactive robots for psychological enrichment. *Proceedings of the IEEE*, 92(11):1749–1758, 2004.

[198] Candace L. Sidner, Christopher Lee, Cory D. Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1):140–164, 2005.

[199] Karine Silva, Mariely Lima, Carla Fafiães, Jorge Sinval, and Liliana de Sousa. Preliminary Test of the Potential of Contact With Dogs to Elicit Spontaneous Imitation in Children and Adults With Severe Autism Spectrum Disorder. *The American Journal of Occupational Therapy*, 74(1):7401205070p1–7401205070p8, 12 2019.

[200] Sara Silva, Filomena Soares, Ana Paula Pereira, Sandra Costa, and Fátima Moreira. Robotic tool to improve skills in children with asd: A preliminary study. *International Journal of Life Science and Medical Research*, 3:162–172, 08 2013.

[201] Vinícius Silva, Filomena Soares, João Sena Esteves, Cristina P. Santos, and Ana Paula Pereira. Fostering emotion recognition in children with autism spectrum disorder. *Multimodal Technologies and Interaction*, 5(10), 2021.

[202] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1145–1153, 2017.

[203] Ellen A. Skinner, Thomas A. Kindermann, and Carrie J. Furrer. A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69(3):493–525, 2009.

[204] I. Smeekens, Manon de Korte, J. Staal, Emilia Barakova, and Jeffrey Glennon. A randomized controlled trial of the effectiveness of pivotal response treatment with and without use of a nao robot in young children with asd. 05 2018.

[205] Filomena O. Soares, Sandra C. Costa, Cristina P. Santos, Ana Paula S. Pereira, Antoine R. Hiolle, and Vinícius Silva. Socio-emotional development in high functioning children with autism spectrum disorders using a humanoid robot. *Interaction Studies*, 20(2):205–233, 2019.

[206] SoftbankRobotics. Nao: User guide, 2022.

[207] Sudha Srinivasan, Inge-Marie Eigsti, Timothy Gifford, and Anjana Bhat. The effects of embodied rhythm and robotic interventions on the spontaneous and responsive verbal communication skills of children with autism spectrum disorder (asd): A further outcome of a pilot randomized controlled trial. *Research in Autism Spectrum Disorders*, 27:73–87, 2016.

[208] Nitish Srivastava and Ruslan Salakhutdinov. Multimodal learning with deep boltzmann machines. *J. Mach. Learn. Res.*, 15(1):2949–2980, Jan 2014.

[209] J. Sung, R.E. Grinter, and H.I. Christensen. Pimp my roomba: designing for personalization. *In Proc. ACM Conference on Human Factors in Computing Systems*, pages 193–196, 2009.

[210] A.R. Taheri, M. Alemi, A. Meghdari, H.R. PourEtemad, and S.L. Holderread. Clinical application of humanoid robots in playing imitation games for autistic children in iran. *Procedia - Social and Behavioral Sciences*, 176:898 – 906, 2015. International Educational Technology Conference, IETC 2014, 3-5 September 2014, Chicago, IL, USA.

[211] Zihan Tan, Anjia Zhou, Xiaojun Hei, Yayu Gao, and Chengwei Zhang. Towards automatic engagement recognition of autistic children in a machine learning approach. In *2019 IEEE International Conference on Engineering, Technology and Education (TALE)*, pages 1–8, 2019.

[212] F. Tanaka and S. Matsuzoe. Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, 1, 2012.

[213] Adriana Tapus, Andreea Peca, Amir Aly, Cristina Pop, Lavinia Jisa, Sebastian Pintea, Alina S Rusu, and Daniel O David. Children with autism social engagement in interaction with nao, an imitative robot: A series of single case experiments. *Interaction studies*, 13(3):315–347, 2012.

[214] Zhansaule Telisheva, Aizada Turarova, Aida Zhanatkyzy, Galiya Abylkasymova, and Anara Sandygulova. Robot-assisted therapy for the severe form of autism: Challenges and recommendations. In *International Conference on Social Robotics*, pages 474–483. Springer, 2019.

[215] Bolat Tleubayev, Zhanel Zhexenova, Aliya Zhakenova, and Anara Sandygulova. Robot-assisted therapy for children with adhd and asd: a pilot study. In *Proceedings of the 2019 2nd International Conference on Service Robotics Technologies*, pages 58–62. ACM, 2019.

[216] Bolat Tleubayev, Zhanel Zhexenova, Aliya Zhakenova, and Anara Sandygulova. Robot-assisted therapy for children with adhd and asd: A pilot study. In *Proceedings of the 2019 2nd International Conference on Service Robotics Technologies*, ICSRT 2019, page 58–62, New York, NY, USA, 2019. Association for Computing Machinery.

[217] Daniel C. Tozadore, Adam H.M. Pinto, and Roseli A.F. Romero. Variation in a humanoid robot behavior to analyse interaction quality in pedagogical sessions with children. In *XIII Latin American Robotics Symposium and IV Brazilian Robotics Symposium (LARS/SBR), Recife, Brazil, 8-12 Oct. 2016*, pages 133–138, Manhattan, New York, U.S., 2016. IEEE.

[218] Renée Van den Heuvel, Monique Lexis, Rianne Jansens, Patrizia Marti, and Luc Witte. Robots supporting play for children with physical disabilities: Exploring the potential of iromec. *Technology and Disability*, 29:109–120, 08 2017.

[219] Caroline van Straten, I. Smeekens, Emilia Barakova, Jeffrey Glennon, Jan Buitelaar, and Aoju Chen. Effects of robots' intonation and bodily appearance on robot-mediated communicative treatment outcomes for children with autism spectrum disorder. *Personal and Ubiquitous Computing*, 22:379, 04 2018.

[220] C.L. van Straten, J. Peter, and R. Kühne. Child-robot relationship formation: A narrative review of empirical research. *Int. J. of Soc. Robotics*, 2019.

[221] Bram Vanderborght, Ramona Simut, Jelle Saldien, Cristina Pop, Alina Rusu, Sebastian Pintea, Dirk Lefeber, and Daniel David. Using the social robot probo as a social story telling agent for children with asd. *Interaction Studies*, 13, 12 2012.

[222] Georgios Velentzas, Theodore Tsitsimis, Inaki Rano, Costas Tzafestas, and Mehdi Khamassi. Adaptive reinforcement learning with active state-specific exploration for engagement maximization during simulated child-robot interaction. *Paladyn, Journal of Behavioral Robotics*, 9:235–253, 08 2018.

[223] J. Wainer, K. Dautenhahn, and B. Robins. Using robots to foster collaboration among groups of children with autism in an after-school class setting: An exploratory study. *Proc. of 1st Workshop on Design for Social Interaction through Physical Play.at the 2nd International conference on Fun and Games*, Eindhoven, The Netherlands, 22-24 October 2008.

[224] J. Wainer, K. Dautenhahn, B. Robins, and F. Amirabdollahian. A pilot study with a novel setup for collaborative play of the humanoid robot kaspar with children with autism. *International Journal of Social Robotics*, 6:45–65, 2014.

[225] Joshua Wainer, Ester Ferrari, Kerstin Dautenhahn, and Ben Robins. The effectiveness of using a robotics class to foster collaboration among groups of children with autism in an exploratory study. *Journal of Personal and Ubiquitous Computing*, 14:445–455, 2010.

[226] Joshua Wainer, Ben Robins, Farshid Amirabdollahian, and Kerstin Dautenhahn. Using the humanoid robot kaspar to autonomously play triadic games and facilitate collaborative play among children with autism. *IEEE Transactions on Autonomous Mental Development*, 6(3):183–199, 2014.

[227] Gabriela Walker and Jennifer Weidenbenner. Social and emotional learning in the age of virtual play: technology, empathy, and learning. *Journal of Research in Innovative Teaching Learning*, 12, 2019.

[228] Ming-Te Wang, John B. Willett, and Jacquelynne S. Eccles. The assessment of school engagement: Examining dimensionality and measurement invariance by gender and race/ethnicity. *Journal of School Psychology*, 49(4):465–480, 2011.

[229] Amy S. Weitlauf, Melissa L. McPheeters, Brittany Peters, Nila Sathe, Rebekah Travis, Rachel Aiello, Edwin Williamson, Jeremy Veenstra-VanderWeele, Shanthi Krishnaswami, Rebecca Jerome, and Zachary Warren. Therapies for children with autism spectrum disorder: Behavioral interventions update, 2014.

[230] I. Werry and Dautenhahn K. Human-robot interaction as a model for autism therapy: An experimental study with children with autism. *In Modeling Biology: Structures, Behaviors, Evolution. Manfred Laubichler and Gerd B. Müller eds., Vienna Series in Theoretical Biology, MIT Press*, pages 283–299, 2007.

[231] Iain Werry, Kerstin Dautenhahn, Bernard Ogden, and William Harwin. Can social interaction skills be taught by a social agent? the role of a robotic mediator in autism therapy. *Cognitive Technology: INSTRUMENTS OF MIND*, 2117, 11 2001.

[232] WHO. International classification of functioning, disability and health. *World Health Organization*, 2001.

[233] Lorna Wing. *The Autistic Spectrum: A Guide for Parents and Professionals.* 1996.

[234] Wisevoter. Autism rates by country, 2023.

[235] Luke J. Wood, Abolfazl Zaraki, Ben Robins, and Kerstin Dautenhahn. Developing kaspar: A humanoid robot for children with autism. *International Journal of Social Robotics*, 2019.

[236] Luke J. Wood, Abolfazl Zaraki, Michael L. Walters, Ori Novanda, Ben Robins, and Kerstin Dautenhahn. The iterative development of the humanoid robot kaspar: An assistive robot for children with autism. In Abderrahmane Kheddar,

Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel, and Hongsheng He, editors, *Social Robotics*, Cham, 2017. Springer International Publishing.

[237] Holly A. Yanco and Jill L. Drury. Classifying human-robot interaction: an updated taxonomy. *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 3:2841–2846 vol.3, 2004.

[238] J.H. Yousif, H.A. Kazem, and M.T. Chaichan. Evaluation implementation of humanoid robot for autistic children: A review. *IJOCAAS*, 6(1):412–420, February 2019.

[239] Woo-han Yun, Dongjin Lee, Chankyu Park, Jaehong Kim, and Junmo Kim. Automatic recognition of children engagement from facial video using convolutional neural networks. *IEEE Transactions on Affective Computing*, 2018.

[240] Aida Zhanatkyzy, Zhansaule Telisheva, Aida Amirova, Nazerke Rakhymbayeva, and Anara Sandygulova. Multi-purposeful activities for robot-assisted autism therapy: What works best for children's social outcomes? In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '23, page 34–43, New York, NY, USA, 2023. Association for Computing Machinery.