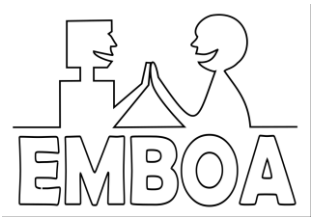


ER-RIA Guidelines

Guidelines for Emotion Recognition in Robot-supported Interventions in Autism

EU Erasmus Plus EMBOA Project

http://emboa.eu/

Version: 1.2, English, July 2022

GUIDELINE AUTHORS

Duygun Erol Barkana
Katrin Bartl-Pokorny
Hatice Kose
Agnieszka Landowska
Michal R. Wrobel
Ben Robins
Tatjana Zorcec

FUNDING

This publication was supported in part by the Erasmus Plus project of European Commission: EMBOA, Affective loop in Socially Assistive Robotics as an intervention tool for children with autism, contract no 2019-1-PL01- KA203-065096. The funding body did not influence the contents of this guidelines in any way.

DISCLAIMER

SUGGESTED CITATION FOR PUBLICATION

<to be added after publishing>

FOR FURTHER INFORMATION ABOUT THE ER-RIA GUIDELINE

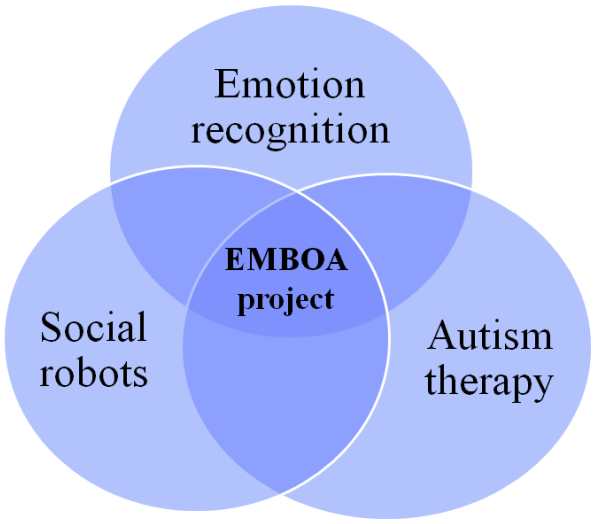
EMBOA project website: http://emboa.eu/
Contact author: Agnieszka Landowska, nailie@pg.edu.pl

List of contents

Summary

The document presents guidelines and practical evaluation of applying emotion recognition technologies in robot-supported intervention in children with autism. The guidelines were developed under the EU Erasmus Plus EMBOA project. The guidelines focus on combining emotion recognition technologies with socially assistive robots. The main goal is to add an affective feedback loop to robot-based intervention in autism therapy.

Visual abstract



Keywords

Socially assistive robots (SAR); autism; autism spectrum disorder (ASD); emotion recognition; affect recognition

1. Context and motivation

1.1. About the document

The document contains a list of guidelines and recommendations for using emotion recognition technology in the observation of the interaction between a robot and a child with autism. The document is prepared under the European Union Erasmus Plus project named EMBOA.

The document is the main output of the project and aims at describing and explanation of the lessons learned via literature review and observational studies. The document is prepared in English, Polish, Macedonian, German, and Turkish. The document is available free of charge under the Creative Commons license CC-BY.

The document in all language versions is available at the EMBOA project website: http://emboa.eu/.

1.2. Motivations

Autism is a lifelong disability that affects people's world perception and interaction with others. It is defined as deficits in social communication and interaction and restricted/repetitive patterns of behaviour/interests/activities and affects 1% of the population (7.5m European citizens). It is a disabling condition with difficulties in independent living, self-care, educational and employment prospects. Almost half of the individuals will have an intellectual impairment and never develop speech. There is no 'cure' for autism. However, there are a range of interventions for enhancing learning and development. Any intervention should focus on developing a child's social skills, as it has been shown that social competence is a predictor of long-term outcomes for individuals with autism. Autism is a highly heterogeneous disorder requiring personalised and tailored interventions for each individual. What may work for one child may not be for the other. Without appropriate intervention, autism can lead to family breakdowns, mental illness, and family members becoming lifelong carers.

Children with autism spectrum disorder (ASD) suffer from multiple deficits, and limited social and imitation skills are among those that influence their ability to be involved in interaction and communication. Limited communication occurs in human-human interaction and affects relations with family members, peers, and therapists. Emotional skills are also among those that could be impaired in children with ASD.

Communication deficits in children with autism are also present in communication between the child and the therapist during the interventions. Recently, the scientific community has been exploring the promising results in robot-based intervention in supporting the social and emotional development of children with autism. Using robots as social mediators to engage children in tasks, allows for predictable and reliable environments e.g. predictable rules is important in promoting prosocial behaviours in autism.

We do not know why children with autism are eager to interact with robots but not with humans. The effect occurs even if the robots have a human-like appearance. Some psychologists attribute this phenomenon to the need for predictability apparent in autism. Regardless of the reason, social robots proved to be a way to get through the social obstacles of a child and make him/her involved in the interaction. Once the interaction happens, we have a unique opportunity to engage a child in gradually building and practising social and emotional skills.

The social interaction with robots might be enhanced with an emotional aspect, adding an affective loop to the process. In order to have an affective loop taking place, emotions of children with autism should not only be influenced but also perceived.

The EMBOA project aimed at performing a feasibility study and proof of concept in using a robot in conjunction with automatic emotion recognition technologies as a novel intervention tool for children with autism.

1.3. About the project

EMBOA project (Affective loop in Socially Assistive Robotics as an intervention tool for children with autism) aimed to develop the guidelines and practical tests for applying emotion recognition technologies in robot-supported intervention in children with autism. The project is financed by the European Union Erasmus Plus Strategic Partnership for Higher Education programme.

The project connects higher education partners specialised in the fields of working with children with disabilities, robotics, and provision of emotion recognition technologies.

Project partners include:

- Gdansk University of Technology, Poland,
- University of Hertfordshire, UK,
- Istanbul Teknik Universitesi, Turkey,
- Yeditepe University, Turkey,
- Macedonian Association for Applied Psychology, North Macedonia,
- University of Augsburg, Germany.

EMBOA Project international meeting on social robots is pictured in figure 2.

Fig. 2. Interaction with social robots, EMBOA project training, Hertfordshire, UK.

The innovation brought by the project is the combination of the two: social robots and automatic emotion recognition in specific interventions in children with autism. The affective loop created in this manner is expected to enhance the set of interventions addressing emotional intelligence skills.

The proposed project is a highly interdisciplinary initiative combining robotics, ICT, and other disciplines like cognitive sciences, developmental psychology, pedagogy, human-machine interface, and others, to open a dedicated possibility for technologies to meet the needs of children with autism and their caregivers. The use of humanoid robots in interventions for children with autism has been growing in the last few years, and the initial research results are very promising. Furthermore, assistive robot technology in interventions for children with autism is innovative by itself, and we also aimed at adding automatic emotion recognition.

The EMBOA project goal was to confirm the possibility of the application (feasibility study) of robot-based intervention and emotion recognition technologies. In particular, we aimed to identify the best practices and obstacles in using the combination of the technologies. The main research question to be answered by the project is: **How to effectively monitor, represent and interpret children's affect by social robots to facilitate emotional states that support the interaction process?**

The main project tasks include:

- State-of-the-art summary of emotion recognition applied in human-robot interaction in autism intervention, based on systematic literature review;

- Performing feasibility study confirming the possibility of the application of emotion recognition in robot-based intervention;
- Identification of the best practices and obstacles in using the combination of the technologies;
- Guidelines develop and practical evaluation of applying emotion recognition technologies in robot-supported intervention in children with autism.

2. Purpose and users of guidelines

2.1. Purpose of the guidelines

Automatic emotion recognition technologies are a relatively new discipline with a growing set of application contexts. All the observation channels used for emotion recognition (facial expressions, the prosody of speech, physiological signals etc.) are characterised by susceptibility to manipulation and disturbances. The level of these disturbances is additionally dependent on the individual and the context. The use of emotion recognition in children with autism is not well explored yet. Using the humanoid robot in conjunction with emotion recognition technologies has rarely been tried and studied. It is a novel approach designed to support and stimulate children with autism in strengthening emotional skills.

The guidelines aim to summarise recommendations and challenges in using automatic emotion technologies in robot-based intervention in autism. We hope to obtain an **affective loop**, i.e. methods for perceiving and influencing emotions of children with autism in robot-based interventions. Accurate emotion recognition in robot-assisted interventions would significantly benefit children and their therapists. Here are some examples: The robot could react in real time to the child's emotional state. This would allow a more natural conversation with the children, preparing them better for real-life conversations with their peers, parents, therapists, etc. Automatic emotion recognition could reduce the amount of manual control of the robot needed via the therapist during robot-supported intervention sessions. This would allow the therapist to concentrate more on the child to maximize the intervention session's success.

2.2. Target group

The target group of the guidelines is diverse, including **therapists and caregivers**, and **researchers** and **technology providers**.

The **therapists and caregivers** might learn about the possibility of enhancing therapy for children with autism with new technologies - social robots, automatic emotion recognition, and a combination of the two.

The **researchers** from both technological and pedagogical domains might benefit from our observations on the previous studies and some recommendations on performing and reporting them.

Technology providers might find interesting the options for enhancing social robots with automatic emotion recognition technologies. In addition, emotion recognition technology providers might benefit from our observations on the limitations using the technologies in perceiving emotions of children with autism.

3. Development method

The development of guidelines is based on several methods. First, we performed a literature review to identify the state-of-the-art robot-based intervention in autism and automatic emotion recognition applied to children with autism. Secondly, we performed two rounds of observational studies on robot-based intervention. Finally, the guidelines were evaluated using the AGREE II instrument. The guidelines were also applied during the second round of observations.

3.1. Systematic literature reviews

Two systematic literature reviews supported the development of the guidelines. The first one concerned the application of automatic emotion recognition on children with autism, while the second considered robot-based intervention in autism therapy. The first study aimed to explore the state-of-the-art in the combination of autism therapy and automatic emotion recognition technologies. To be more precise, in the field of interest, some studies show how to automatically recognize emotions felt by children with autism, not the capacity of children to recognize emotion in others. Over 2000 papers were initially extracted from 7 search engines, including 50 papers in qualitative and 27 in quantitative analysis. The study reveals some observations regarding observation channels, modalities, and methods used for emotion recognition in children with autism. Qualitative analysis revealed important clues on participant group construction and the most common combinations of modalities and methods. The study might interest researchers who apply emotion recognition or enhance methods for affect classification in autism-related studies. This systematic literature review revealed a number of challenges related to applying f emotion recognition to studies on children on the autism spectrum. Some good practices were also identified. The study used the PRISMA (Preferred Reporting Items for Systematic Reviews) execution and reporting scheme and was published. [2]

The second study aimed to explore the state-of-the-art use of social robots in intervention for children with autism. Existing literature suggests that children with ASD benefit from robot-based interventions. However, studies varied considerably in participant characteristics, applied robots, and trained skills. Therefore, robots might be divided into five categories based on their morphological characteristics: humanoid, animal/creature, mobile robot, ball-shaped robot, and others. Among those, the most important role in autism therapy is assigned to robots that aim to develop interaction in children (called social robots). Some social robots are depicted in Figure 1. Three are humanoid, while the last one is an animal (puppet).

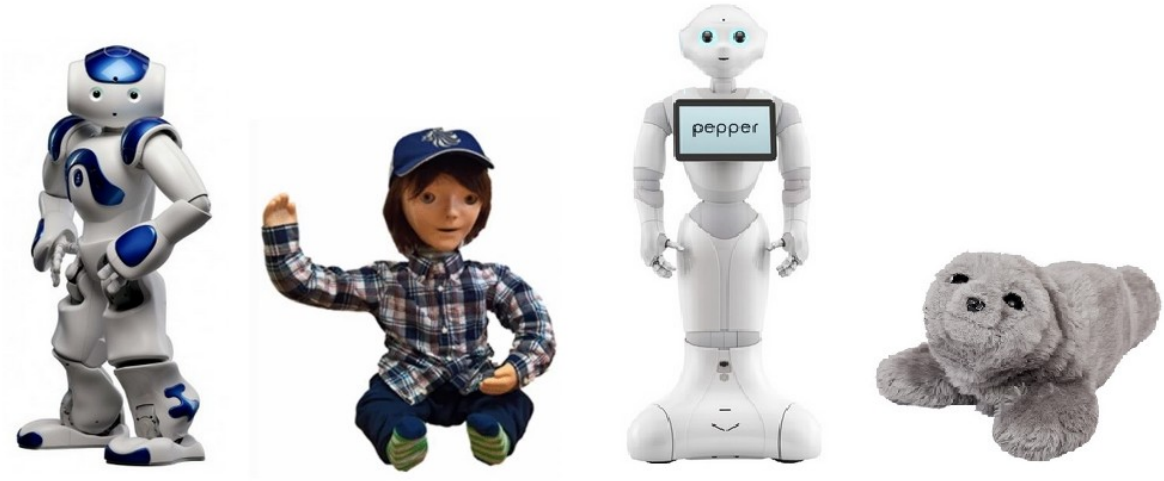


Fig. 1. Examples of social robots (left to right): Nao, Kaspar, Pepper, Paro

We reviewed papers describing robot-based interventions targeting diverse skills for children with ASD systematically retrieving all relevant articles published using the databases Scopus, Web of Science, and PubMed. From a total of 609 identified papers, 60 publications, including 50 original articles and 10 non-empirical articles including review articles and theoretical articles, were eligible for the synthesis. Nao and ZECA were the most frequently used robots. Recognition of basic emotions and getting into interaction were the most frequently trained skills, while happiness, sadness, fear, and anger were the most frequently trained emotions. The studies reported a wide range of challenges with respect to robot-based intervention, ranging from limitations for certain ASD subgroups and security aspects of the robots to efforts regarding the automatic recognition of the children's emotional state by the robotic systems. Finally, we summarised and discussed recommendations regarding the application of robot-based interventions for children with ASD. The study also used the PRISMA execution and reporting scheme and was published. [1]

3.2. Observational sessions

The development of the guidelines benefits from two rounds of observational studies where emotion observation technologies were applied. The studies were held within EMBOA project in Macedonia, Poland, United Kingdom, and Turkey. Common interaction scenarios and protocols were used to perform the studies. The first round of observational studies was prepared according to findings from literature reviews. Then we have reviewed the data from the study, and formulated the guidelines accordingly. Then the second round of studies was performed with some modifications according to the lessons learned.

Some general assumptions for observation sessions were as follows:

- all setups should be ready before a child enters the room;
- multiple sessions might be held with a child (at least 2);
- familiarisation sessions are encouraged;
- there is a pre-prepared common list of scenarios to perform;

- the scenarios are to be translated into national languages and adjusted, if necessary (for example children sing different songs in different countries);
- we try to follow the specified scenarios, but it was allowed to follow a child - dropping or adding other interactions in between on the run;
- we record video (with 2 cameras, if possible), eye-gaze, voice (with 2 microphones, lapel and general one), physiological signals (heart rate and skin conductance),
- during the session we write down the most important observations, and session might be annotated afterwards;
- we record, store and share data (anonymized and coded) within consortium.

The list of prepared scenarios was as follows:

- Scenario 1. Greetings and a song ("Happy and you know it" or equivalent) - this scenario contains basic greetings (hello, bye, how are you) and a song divided into verses; might be used for training greeting skills, singing together;
- Scenario 2. Emotions scenario - contains showing happy/sad/hiding/surprise pose along with a voice communicate, but also prompts for a child showing similar pose; might be used for training movement imitation, emotion recognition skills;
- Scenario 3. Animals scenario - contains prompts for making animal sounds and the sounds; might be used for training vocalisations, sound imitation, turn-taking, question answering;
- Scenario 4. Body parts scenario - contains poses of Kaspar showing body parts along with voice communicates; might be used for training imitation of movements, showing body parts (Kaspar's or own), turn-taking;
- Scenario 5. Can you copy Kaspar - contains diverse poses of Kaspar (hand to the side, forward, up, etc.) and prompts to copy the movements - might be used for training imitation of movements, turn-taking, memory game;
- Scenario 6. Imitation of sounds - contains selected letters (vowels), syllabes and simple words with prompts for the child to repeat; might be used for training vocalisation, basic speech skills, imitation of sounds, turn-taking, and memory game;

Additionally each of the scenarios contained basic prompts and reinforcements: well done, bravo, thank you, trumpet + bravo, it hurts, your turn. Those are used to encourage a child to follow scenario and interaction with Kaspar. We started session with "Hello" and ended with "Bye". The order of the scenarios might be adjusted.

All of the partners in four countries performing the studies used the same set of equipment, as agreed in the project, for conformance of observations. The equipment we used in our observation studies is given as follows:

- Affectiva E4 wristband for capturing physiological signals;
- Gazepoint GP3 Eye tracker to capture eye-gaze;
- 2 microphones: Zoom H4n Pro and AKG C 417 L lapel microphone with an adapter;
- 1 camera (1st round) and 2 cameras (2nd round).

The particular devices were selected at the project proposal stage and some of those turned out to be good choices. Please note, that the guidelines do not suggest usage of any of those

devices, but we rather provide them for the future reference (and perhaps replication) of the study. You might find interesting the criteria we have used for device selection, and those were as follows:

- the lowest intrusiveness of measurement for a child;
- a possibility of long-term measurements;
- robustness to disturbance;
- data export function;
- quality to price ratio.

During the second round of the studies we have applied some lessons learned (guidelines) and we have made some modifications to the observational sessions:

- using 2 cameras (frontal one close to a child) and general one (scene view);
- camera position and angle adjustment and reporting both resolution and angle of the camera lens;
- using (if possible) alternative cameras (selected partners);
- elimination of eye-tracker resulting in Kaspar sitting lower - and trying to obtain gaze data from frontal RGB video image;
- getting rid of lapel microphone and using general one only (as no better results were obtained from the lapel one);
- recording samples of child's voice, room noise, robot noise and speech;
- adjusting procedure for synchronization of signals.

A total of 65 children with ASD took part in the study, aged 4 to 12. We have also noted: developmental age, type and length of therapy so far, comorbidity, having siblings, rating of: language understanding, verbal skills, animal knowledge, body parts knowledge.

4. Guidelines

The recommendations that we have obtained from the literature reviews and the lessons learned from the observation sessions we conducted were then integrated and summarised in the form of guidelines.

The guidelines are divided into the following categories in order to systematise and make them more searchable:

1. Guidelines regarding the choice of observational channels (denoted as CH)
2. Guidelines regarding the reliable acquisition of emotional symptoms (SYM)
3. Guidelines regarding technologies and devices used (TECH)
4. Guidelines regarding activities in the interaction with the social robot that enabled the proper acquisition of symptoms as well as obstacles (INT)
5. Guidelines regarding symptoms data processing (PROC)
6. Guidelines regarding recognition of emotions in children with autism (EMO)
7. Guidelines regarding research design on the topic (RES)
8. Guidelines regarding reporting the studies on emotion recognition, robots, and children with autism (REP)

The guidelines are numbered for easier reference, and the category code is also assigned for easier identification. In addition to categorised observations, there are 3 proposed general guidelines (GEN) to start with.

Guideline GEN1:
Follow the child and therapeutic goal

Robot-based intervention, as well as obtaining signals for emotion recognition, requires technological appliances. This makes the intervention environment more complex, with appliances requiring some attention from the operator. This guideline means to keep the focus on a child and the needs of the therapy. **It is not the child nor therapy that should be adjusted to fit into emotion recognition technology requirements. Instead, the emotion recognition setup should follow the child's therapeutic purpose.**

Analysing the variety of studies in the development of real-life applications of emotion recognition in autism therapy, some questions seem of the highest importance to be asked at the very beginning of each study:

- For what purpose are emotions recognized - is it to better understand the phenomena of emotions, to support intervention, or to adjust technology (robot, app) behaviour?
- In what way would emotion recognition help to develop skills or support play in children with autism?
- Training of which skill would require automatic emotion recognition? [1]

Guideline GEN2:
Start with what you want to know about child's emotional states

At first, you need to consider what you want to track and what you want to know about the child's emotions. Define which emotional states of a child are of interest, then choose and adjust technology accordingly. Which emotions do we need to detect and for what? It might be easier to detect just selected emotional states.

In psychological research, there is no unambiguous definition of human emotion [28]. However, a discrete emotions concept is widely accepted. There are multiple models to represent emotions [29]. Analysing papers that deal with emotion recognition, we observed the issue of which emotions are distinguished and analysed. Most of the papers use two emotion models: Ekman's basic emotions (joy, anger, fear, disgust, sadness, surprise) [30] and/or a two-dimensional model of valence and arousal. Ekman's basic model papers use a subset of it, not the complete set of emotions. Papers use different wording to describe the emotions, and additional attention should be paid to the meaning of those in a particular study. The set of emotions resulting from the recognition process is not limited to the ones described by these two models. Some studies, apart from emotions, also considered moods, mental states (such as concentration), or even hunger. We additionally analysed other emotional states addressed and the co-occurrence of states - it will not be reported here in detail; please refer to our paper [2].

Sometimes the six basic emotions (happy, scared, angry, surprised, disgusted, sad) are good ones to start with, but they might not be sufficient for study or intervention. For example, some studies reported engagement and early symptoms of stress that were the most important to track [31, 32].

Guideline GEN3:
Protect the child's rights

No matter the purpose of the intervention and/or the study, please be reminded that the child's rights must be protected. First of all, the child has the right to get the best therapy available.

Another issue is to protect the privacy of a child. According to GPDR (https://gdpr-info.eu) personal data is any information related to an identified or identifiable natural person. Identifiers such as a name, identification numbers or address, should not be exposed. Informed consent must be obtained from the caregivers with a separate consent on publication of the results, vocalisations, or facial images. Other ethical risks should be considered as well. It is important to allow the child to refrain from the interaction at any point. Technical risks that compromise safety must be considered as well (see TECH1). Ethical Committee approval should be sought whenever necessary.

4.1. The choice of observational channels (CH)

The nervous system responses of the human organism evoked by emotions cause changes in life activities that generate modalities such as facial expression, body posture, vocalisation, heart rate, skin conductance or peripheral temperature. This observable information, which can be used to obtain insight of emotion symptoms, can be obtained from multiple channels. A **channel** is a medium for recording a signal, eg. video, voice. **Modality** is a type of information observable and used as a proxy for emotion recognition. In previous studies, modalities used for emotion recognition, were as follows.

- movement: facial expressions, body postures, eye gaze, head movement, gestures (also called hand movements), and any other not previously classified motion;
- sound expressions: vocalisations, the prosody of speech;
- heart activity: heart rate, HRV (heart rate variability);
- muscles activity not related to movement: muscles tension;
- perspiration: skin conductance;
- respiration: intensity and period;
- thermal regulation: peripheral temperature;
- brain activity: neural activity. [2]

Guideline CH1:
While choosing observational channel consider type of activity, child condition, and context

The usefulness of observation channels for emotion recognition is tightly linked to observed activities.

Therefore, when choosing channels, it is important to consider the type of interaction planned between the child and the robot. For example, using an RGB camera during an interaction involving a child's movement, such as physical exercise, may not provide sufficient data for emotion recognition, due to stepping outside the camera's field of view or not capturing the child's face.

Another circumstance to consider is the child's condition. Some children with autism spectrum disorder may refuse physical contact, preventing the attachment of physiological recording devices such as a wristband. Consider a child's speech skills and level of functioning as well.

The choice of observation channels should take into account also the context (incl. place, timing) of the study being conducted.

Guideline CH2:
Observational channels may not provide useful data during the entire observation.

Whichever channel is used, some data may be periodically missing. The reasons for such a situation would differ depending on the channels or devices used. For example, when capturing images with an RGB camera, the subject may move out of the field of view, look down or sideways, or have face partially covered. When collecting physiological data with a wristband, rapid hand movements or fidgeting may cause erroneous readings. These are situations that cannot be avoided, and they are inherent when recording children.

You need to be aware of the risks and handle data processing adequately. It is advised to monitor quality of data provided by a specific channel over time and remove time windows when symptoms are not clearly visible. You might consider multimodal observation (see CH3) or multiplication of a single modality, for example multiple cameras (see SYM2) to gain more availability.

Guideline CH3:
Consider multimodal observation for reliability and availability

As any channel might be temporarily unavailable (guideline CH1), a natural idea appears to use multiple channels or devices.

One might distinguish multi-channel and multi-model observation. The first one refers to the number of devices used for capturing observable symptoms, eg. using a combination of a camera and a microphone, but also using multiple cameras. The second one refers to analysing diverse symptoms e.g. facial expressions, prosody of speech. Please note, that multi-modal observation (e.g. facial expressions and prosody of speech) might be performed based on a single channel (both modalities retrieved from a single camera) or multiple channels (using both camera and microphone).

Multimodal (or multichannel) analysis of affect provides availability and reliability. Regarding availability - the more devices you use, the more probable it is that at least one of them would provide valuable information on emotional symptoms. Moreover, multimodal emotion recognition systems offer higher classification accuracy than single-modality-based solutions [6]. Therefore, if feasible, it is advisable to use multiple channels during observations to collect different modalities.

Once the multimodal or multichannel observation is applied, there are some other challenges regarding the data fusion from channels (please see guidelines PROC1 and PROC2).

Guideline CH4:
Limit number of devices

Children with autism usually like repeatable, same environments. All intrusions, including devices put on the body, devices in the room, and extra people, might cause a

refusal to participate or might be a reason to behave differently. For example, a study on smile recognition utilises wireless EMG put on a child's face. However, the study reports that 70% of children agreed to wear facial EMG devices [33], which means that some children could simply not obtain the data. In addition, the study did not measure the influence of wearing the device on child behaviour.

The more devices you will use, the more likely it is that a child would refuse to participate due to the novelty and complexity of the environment (cameras, computers, cables, extra people in the room). We are well aware that this guideline opposes the previous one (CH2). There is a trade-off between obtaining reliability and availability versus simplicity of the environment.

One might consider using a single device to capture multiple modalities. For example, RGB video enables the data collection on facial expressions, body posture, or gestures [2]. In addition, you might consider obtaining eye gaze from the video instead of using costly and hard-to-calibrate eye trackers. Sound could also be obtained from a high-quality video camera to avoid potentially distracting microphones on the table in front of the child or on the child's clothes.

An example of a minimal set of devices that would provide multimodal data might be an RGB camera and a biosignal recording wristband. Using the camera, it is possible to obtain modalities related to movement (e.g. facial expressions, eye gaze, head movement) and sound (e.g. prosody of speech, vocalisation). On the other hand, various biosignals (e.g. heart rate, skin conductance, temperature) can be recorded using wristbands [2].

4.2. Reliable acquisition of emotional symptoms (SYM)

The second challenge that we have identified is that even though one will record the entire session using devices, the data will not contain the sufficient information to conclude symptoms of emotions. This applies for example to eye gaze tracking, which is hard to calibrate, but also to other modalities. In addition, the acquisition of emotional symptoms is prone to multiple confounding factors that might influence the reliability of the observation. Therefore it is worth paying attention to conditions of symptoms recording, and those conditions apply to specific types of modalities obtained.

Guidelines for child's facial expressions acquisition

Guideline SYM1:
Adjust the distance between a child face and a camera to the resolution of the camera.

In order to recognize emotions from facial expressions the whole child's face must be visible as a fraction of the obtained video image. For the optimal video frame should capture a child's face as a minimum 10% fraction of the total image. For example, when using a typical small internet camera, the child's face should be around 1-1.5m away from the camera's lens. The distance might be adjusted depending on camera resolution and angle. Telphoto lenses-camera might be considered as well. Otherwise, the effectiveness of emotion recognition based on facial expression decreases.

As already mentioned before, some automatic emotion recognition solutions require the face to be at least a certain fraction of the total scene image. However, it is possible to cut and enlarge the face in post-processing, however, this is time-consuming and might result in low image quality when applied to low-resolution video. Instead, position the camera close to the child's face and check the camera resolution settings to get the optimal recording.

On the other hand, the camera distance might be too small when the angle that the camera covers is small. In such conditions, a child moving forward or sideways (which is impossible to eliminate) might record only a fraction of a face, which also hinders automatic emotion recognition results. Using ultra-wide angle cameras might be an option to address this challenge.

Check the visibility of the whole face and size of the child's face with regard to total image before starting the recording to get better recognition results.

Guideline SYM2:
Use more than one camera and position at least one camera right in front of a child

Although in guideline CH3, we advise using a single device to capture multiple modalities, one must be warned that observation with a single camera also is prone to some risks. For example, when a child is moving around the room or even turning sideways while being seated, a single camera might not be enough to capture the symptoms during the entire observation. Consider putting different cameras in different locations to extend the time span of capturing the child's face from the right angle. A single camera located too high or sideways will give you a scene recording but not sufficient face image quality. Angular or tilted views might make facial expressions hard to analyse as well. When using more than one camera pay attention to synchronisation challenge (see PROC1 and PROC2).

The effectiveness of emotion recognition based on facial expression analysis is significantly affected by the position of the camera relative to the face [7]. For example, positioning the camera high above the face causes recognition biases toward the emotion of anger and too low location biases toward surprise. A similar bias can occur with vertical alignment. Therefore, especially if only one camera is used, it should be placed as frontal as possible.

If multiple cameras are used, additional side cameras can capture behavioural patterns such as posture and gestures.

For better automatic emotion recognition results, limit the presence of other people in the camera's field of view.

Guideline SYM3:
Masks, glasses, and other face occlusions make it more difficult or sometimes impossible to obtain reliable facial expressions for affect analysis.

Facial occlusions can significantly reduce the effectiveness of emotion recognition based on facial expression analysis. The most common include glasses, especially the ones with thick frames. Hair cuts with long fringes that cover the eyes also influence recognition of the face and thus emotional expressions. Due to the pandemic, masks were recently added to the common face occlusions list.

Whenever possible, try to reduce occlusions, e.g. by removing the mask in a safe environment or moving the hair. However, in some cases this will not be an option (glasses, sometimes mask). Therefore, it is important to report this situation to take into account the likely low accuracy of emotion recognition in this modality when processing the collected data.

Guideline SYM4:
Adjust the illumination level of a child's face.

It is important to ensure proper lighting conditions when recording video. Excessive illumination will cause overexposure, resulting in a loss of image detail, including, for example, facial features, which are key to recognizing emotions based on an expression. On the other hand, poor lighting will cause underexposure, which involves the loss of shadows that also allow detection of facial features. Overexposure, underexposure, and uneven illumination have highly undesirable effects that might significantly reduce the effectiveness of emotion recognition methods based on facial expressions.

Before starting a session with children, check the lighting conditions by recording a test video. Adjust the lighting to capture facial details such as the nose, eyebrows, and mouth well. Please pay attention to the location of the windows (which are not good as a background and might cause uneven illumination of the face if a child is seated sideways).

Guidelines for child's vocalisation acquisition

Guideline SYM5:
Choose a silent room - control the level of the outside noise and limit the noise generated by the room interior.

In order to obtain high-quality voice data of the children, the intervention sessions should take place in a quiet environment with only minor background noise. For example noise from nearby streets, playgrounds, etc. should be avoided.

It might help to control the outside noise level if you place an information sign in front of the room asking for silence because a study/therapy session is currently taking place. If possible, please make sure to close the windows when you start the data collection.

If possible, do not choose a room with a long reverberation time (echo) for your study/therapy. However, it can help to place curtains in front of the window, add noise-absorbing furniture, and/or decorate the room with objects such as soft toys or pillows (but not too much, as they could be distracting). The latter may also help the child to feel more comfortable.

Try to avoid as much sound as possible originating from chairs, tables, doors, furniture, and windows as possible. Here are some examples of how to do that: Close the window before the data collection starts, place all needed toys etc., next to you on a table/in a bag/etc, so that you do not have to stand up and get these objects out of a cupboard during the intervention session. Place rugs beneath the table, chairs, and feet of all people in the room.

It is also recommended to ask the therapist and the parents to produce as little noise as possible with their shoes, hands, and used objects.

Take care of background noise when you position the microphone: It is not recommended to position the microphone on tables or clothes that produce a loud noise when touched or moved, as this may lower the detected child vocalisation events.

Guideline SYM6:
Limit the co-occurrence of the child's and the other's voice.

Emotion recognition based on audio signal benefits from voice activity detection [9]. However, the child's voice detection is hampered if the child's voice co-occurs with other people's voices [10]. Therefore, we recommend reducing the number of people present in the room during the intervention session. It is likely not possible to reduce the amount of speech produced by the therapist, but you may consider asking accompanying parents to try not to talk simultaneously with the child during the intervention. Moreover, the therapist and the parents should avoid talking simultaneously and should be encouraged to postpone unnecessary conversations with each other, i.e., conversations that are not related to the course of the current intervention session, to the end of the intervention session.

It is recommended to note the number of adults present during an intervention session for study purposes.

Guideline SYM7:
Adjust location of the microphone to the position of a child and other people present.

To record high-quality child voice data, it is important to position the microphone closer to the child than to other people and the robot in the room as this helps to increase the volume of the child's voice in comparison to other voices. The therapist should sit as far away from the microphone as possible. If the child is talking very softly, you might encourage the child to talk louder so that the robot can process the child's voice.

4.3. Technologies and devices (TECH)

We found it quite challenging to choose, learn, and configure devices we used for recording emotion recognition symptoms during our observational studies. Therefore, those guidelines are quite specific regarding the devices.

Guideline TECH1:
Provide a safe and non-distracting environment for a child.

The observation environment should be safe, first of all for the child, but also for the robot and the recording devices. You should arrange the environment in advance. The robot should be situated stably and securely so that interaction with the child does not cause it to fall over uncontrollably. You should also pay attention to the wires of the connected devices. If possible, do not place them next to children to avoid tripping as well as destroying the measurement environment by pulling the cable. For children with ASD, events such as the fall of a robot or measuring device can cause a severe reaction.

Apart from safety, child's distraction is also a point to consider. Therefore, exposure to unnecessary elements of the measurement environment should be avoided. Computers and recorders can be hidden behind a panel, curtain, or even in a cardboard box, and cables can be duct-taped to the floor. Carpets might also help to hide cables on the floor.

Guideline TECH2:
Select devices based on comfort for the child and device setup effort

Some tools or devices require a complex setup procedure to record and process data properly. Moreover, some of them might be disturbing for a child (see TECH4).

It is possible that voice activity detection does not work well for all types of microphones. Check the performance of the voice activity detection, especially if you plan to use microphones that may disturb children with ASD, such as lapel microphones. In our observation round 1, the lapel microphone did not add more information on vocalisations than the internal microphone [10] so we decided not to use it during the second round as it was disturbing for a child.

Another example is an eye tracker requiring a child to stand/sit still and focus on a series of points displayed on a screen in a timely manner. As such a calibration procedure is hard to follow for a child and even more difficult for a child with autism. The advice is to do one of the following: (1) find a device (eye-tracker) that does not require a calibration procedure or is robust to some faults in calibration; (2) try to experiment with a calibration procedure to simplify it for a child; (3) use an eye-tracker with elder and higher -functioning children only.

Please note that some eye trackers (but not all of them) have a narrow field of recorded and analysed views. In addition, they might require a child not to move around (forward/backward or sideways), which is hard to obtain. Therefore, it is always advisable to experiment with the device, and test it in a practical setting before starting a session with a child. It is hard to calibrate the device and for most of the children the data obtained was hard to analyse during our observation studies.

Consider an alternative approach to obtaining a certain modality - for example, algorithms obtain some eye-gaze data from videos [34].

Guideline TECH3:
For devices that must be put on a child (on hands, attached to clothing), consider familiarisation stage

Take into account that some children may feel uncomfortable with some of the devices, especially if they need to be worn on their body or clothes. Therefore, try to find devices that do not require to be placed on a child's body or clothes. If it is not possible, a familiarisation process that could be individualised depending on the child's specific needs is advisable.

For example, it can be tricky to accept the lapel microphone on their clothes for many children with ASD. Therefore, carefully evaluate if such a microphone is needed to achieve your study/therapy goals. If so, try to find ways to familiarise the children with the microphone. For example, it might help to have similar microphones (or a dummy) for the robot, the therapist, and the caregiver. Some children might be suggested that the robot can only hear them if they wear this microphone.

For the devices that capture physiological signals (such as the E4 wristband that we used in our studies) choose something cable-free. Wristbands are a good option as they seem like a watch and might get accepted by the child more easily. In addition, it could help some children to feel comfortable if the researcher/therapist or caregiver first shows the device to

them, lets them press the button, etc. It might help to wrap the wristband in a gentle coloured cloth for other children. Another possibility is to assemble similar wristbands (or dummies) on Kaspar, the therapist, and the caregiver, so that wristbands become something special to that specific context.

Be aware that children tend to manipulate devices that they can reach with their hands (e.g., wristband, lapel microphone). This is even more the case if they see at the beginning how to turn the device on or off. Therefore, only show the child how a device works if no other way of familiarisation helps.

Guideline TECH4:
Adjust recording levels of the microphone so that the child's voice is well captured

If the audio recording levels of the microphone are low, the child voice activity detection may not work properly resulting in no or only a few detected child vocalisations. Please carefully select the settings before the start of the observation session. Adjust the microphone configuration to the room, background noise, microphone location, and child's vocalisations volume. It is advisable to perform a test session before the observation sessions (without a child).

4.4.Activities in social robot interaction (INT)

There are promising results of using robots to support the social and emotional development of children with ASD. Robots as social mediators for engaging children in tasks allows for a predictable and reliable environment; e.g., having predictable rules is an important prerequisite in promoting prosocial behaviours. We do not know exactly why children with ASD are eager to interact with human-like looking robots and not with humans. Regardless of the reason, social robots proved to be a way to get through the social obstacles of a child and make him/her involved in the interaction. Once the interaction happens, we have a unique opportunity to engage a child in gradually building and practising social and emotional skills. [1]

Guideline INT1:
Adjust emotion recognition goals and technologies to the intervention or therapy

We analysed which skills were taught in robot-based interventions in previous studies. We provide a list here for your reference - it might give you an idea of what social robots might be used for [1]:

- social convention (greeting skill, singing together, sharing, closing interaction, ability to thank, saying please);

- social interaction (getting into interaction, turn-taking, following / imitating movements, social attention abilities, child's engagement in an activity, sensory processing skills, basic eye contact, conversational interaction, socio-emotional behaviours, initiating, focus on self-initiated interaction, playing a game together, requesting an object);
- social responding (eye gaze following, response to a behavioural request, response to name)
- emotional skills (recognition of basic emotions, mapping emotions and sounds, context-emotion association, discrimination between thoughts and emotions, reading emotions, mimicking emotions learned in the robot training);
- control (touching to transfer the robot to a positive emotional state, adaptive behaviours in situations associated with anger and sadness, control of negative emotions in social situations);
- other skills (self-care, cognitive skills, improve learning, drumming game, rhythmic upper and lower body interpersonal synchrony, moving on a count, moving on a steady beat, selecting particular colours).

Carefully think about how the intervention and the robot-child-interaction can benefit from emotion recognition. Therefore, it is recommended to specify the goals of emotion recognition for a given child and therapy goal. An accurate emotion recognition is not always needed.

The automatic emotion recognition technologies are especially important when emotional skills are trained. The set of emotions recognized might be adjusted to the training purpose and the child's ability. For example, during the smile training, only "happy" expression might be tracked.

On the other hand, it might be interesting to capture the general mood or attitude of a child during any of the interventions. For example, one might distinguish fear, anger, boredom, or engagement among the states of interest.

In some cases, it might be sufficient to automatically recognize if a child is crying or not, or just distinguish positive and negative reactions.

Guideline INT2:
Plan familiarisation stage for a child to get used to a robot and other devices present.

For most young children, especially if they had no prior experience with robot-supported intervention, it can be useful to have a familiarisation session with the robot and all devices before recording of the study data (especially devices put on children - see TECH4). Some children might require familiarisation with the unknown room and people.

It is important to allow the child to refrain from the interaction at any point (see GEN3). The therapist/researcher should encourage and motivate the child to interact with the robot but not press the child.

Guideline INT3:
Plan activities within the session in advance, but adapt the plan on the run, whenever necessary.

The intervention should follow the child, i.e., it is good to have a vague plan of the intervention in advance, but researchers/therapists should be flexible enough to adapt the intervention activities to the child's current needs.

At the beginning of the robot-supported intervention, using ice breakers, such as the robot introducing himself or motivating the child to sing a funny song together is recommended. It might be good to have different kinds of ice breakers that should fit the individual child's age, interests, and verbal skills.

The intervention itself should be taken into account that ASD is a very heterogeneous disorder. Therefore, the intervention should be adjusted to the individual child's attention span and to other characteristics, including verbal skills, cognitive skills, potential comorbidities, age, and preferences. For example, this could be done by adapting the length of the session, the choice, the order and the length of the activities, and the number and length of breaks.

Guideline INT4:
Consider diverse activity types and diverse difficulty levels

It is recommended to prepare different activities for specific ages, verbal and cognitive capacities, attention span, individual interests, and therapeutic goals. Moreover, it is useful to prepare activities for diverse difficulty levels to provide the best suitable intervention to each child with ASD. In order to quickly select the appropriate activities, it is helpful to annotate them according to their difficulty level, for instance.

When selecting the activities for an intervention, the interests of the children should be taken into account. Many children, for example, seem to enjoy singing the most. Therefore, singing a song could be chosen as an ice breaker at the beginning of the intervention session and/or in between to relax the child whenever he or she seems to be stressed. On the other hand, it might be the case that singing a song is not helpful for all children of all ages. Therefore, it would be helpful to find out the interests of a specific child prior to the actual robot-supported intervention session to prepare activities for the child that he or she will likely enjoy. An example for verbal and high-functioning children might be a vowel memo game.

Guideline INT5:
Perform multiple sessions with a child

In order to evaluate the success of the robot-supported intervention in comparison to standard therapeutic approaches, it is highly recommended to follow the child over longer periods of time, i.e., to perform multiple robot interaction sessions with a child. It may be challenging to recruit children performing in longitudinal studies for study purposes. In this case, it could be helpful to embed these robot-supported sessions in the children's regular therapy plan and/or to motivate the families with "goodies", such as participant allowance or toys for the child.

4.5. Symptoms data processing (PROC)

From the literature review, we obtained some observations on the challenges encountered in the studies regarding processing the modalities. The challenges encountered in the studies fall into three categories: (1) the data obtained are not of high quality (hold no meaningful information on the child's symptoms of emotions), (2) the obtained modalities are hard to analyse due to atypical patterns of symptoms, (3) observed symptoms of emotions are contradictory.

Guideline PROC1:
If recording multiple modalities, pay attention to synchronisation

Multimodal emotion recognition systems offer higher classification accuracy than single modality based solutions [6]. The most popular channels include: RGB video, depth video, sound and physiological signals. However, if different devices are used to retrieve data from the channels, problems with synchronisation may occur. In addition, it is very challenging to start recording across all devices at the same time.

Therefore, it is important to elaborate and test the synchronisation strategy in advance. For example, you can note down the exact time each device started recording. Or think of a method like a timestamp (perhaps the device has such a function). You can also, for audio-visual channels, use a film clapper board or a program to start all the devices at the same time. Whichever method you choose, it's important that it must enable channel synchronisation when pre-processing the collected data.

Guideline PROC2:
If recording multiple modalities, pay attention to inconsistencies

When using multi-modal or multi-channel emotion recognition systems (see CH3 and REP2), attention should be paid to the inconsistent results from the individual modalities.

A list of challenges in multimodal integration includes, among others, the integration of the results obtained in differing emotion representation models and dealing with the inconsistency of the results.

We observed inconsistencies in recognition results in the experiments, both based on diverse and the same input channel analysis. [12] For example, in an experiment on an educational management game, a significant discrepancy between the self-evaluated and detected emotional states was reported [13]. In another study we observed differences in the recognized emotional state based on facial expressions recorded with two cameras: placed below and above the monitor [7].

Analysing the same moment of observation with the same person and context, diverse algorithms might provide different results of emotion recognition.

The recommended approach is to evaluate the temporal unavailability of channels as well as data quality before jumping to data processing and analysis. Another recommendation we can give regarding the issue of inconsistent emotion recognition results is to pay attention during interpretation and report any inconsistencies encountered.

Guideline PROC3:
Annotate events on the run (if possible) or in post-processing

During a session of child-robot interaction, diverse events might occur that might influence the data recorded and then result in misinterpretation. Therefore it is crucial to be quite scrupulous in noting those events down. The notes should include interruptions, such as room entries by a third person, and extra noise from outside the room (or inside the room - e.g. sth has fallen). Note down also atypical behaviours of a child and events related to recording devices, such as a dead battery, slip of a sensor, or a child taking a sensor off.

If possible, the best option is to make those notes in real time (sometimes it requires an additional person as an observer), although it might be hard to note all events down. Post-hoc annotation of the recordings (if you have recorded a general scene with a video) is also possible but time-consuming.

Post-hoc annotation is crucial for interpretation of the results, and might be valuable from the perspective of publishing a dataset for future research (see RES2). In order to properly annotate, consider precise definition of events, states and tasks to annotate, annotation by multiple people (including consistency check), and repeatable method (tools) for the process.

Guideline PROC4:
For the voice channel, the child's speech must be distinguished from other sounds first.

In order to recognize the child's emotional state from his or her voice, we first need to determine the segments with vocalisations of the child. Therefore, it is important to distinguish the child's voice from other voices and background noise. We observed that segments that were correctly classified as containing the child's voice mostly consisted of loud and clearly articulated child's speech with no or little background noise [10]. In case of unavoidable background noise - such as noise from nearby waiting rooms - it is helpful if the child's voice is recorded considerably louder than the background noise. Furthermore, it is useful to provide a wide range of training material for the voice activity detector. If the training material does not include a sufficient amount of, for example, singing, imitation of animal sounds, and crying events, it is more likely that the voice activity detector struggles to detect such events as a child's voice [10]. It should also be considered that children of different ages have different voice characteristics.

When planning a study, take into account that a careful evaluation of the performance of a voice activity detector affords the time-consuming tagging of all occurrences of sound types such as child's voice, voices of other people, sounds produced by the robot, and background noise. Consider recording a baseline for a voice channel (see PROC5).

Guideline PROC5:
For some signals, the child's baseline should be recorded.

Recording a baseline might be a good option for further data processing and analysis. A baseline is a fixed point of reference that is used for comparison purposes. For example, for a voice channel, a sample of an undisturbed child's voice might be recorded. For physiological signals a baseline is a recording during a resting phase (preferably: in the other room than the one with a robot).

Humans differ significantly in terms of nervous system reactions to emotions. For example, there are high-, mid- and low-reactivity individuals. Physiological reactions also change with age. Moreover, the group of children with autism is even more divergent due to a unique combination of deficits. The way to consider the diversity while processing the signals is to record an individual's baseline (a physiological response while resting).

Previous studies revealed that emotion recognition baselines for children with autism have different levels than for typically developing children. For example, the paper [16] describes important observations concerning sensors and technologies that can be used in automatic emotion recognition: (1) children with ASD had a significantly lower amplitude of respiratory sinus arrhythmia and faster heart rate than typically developing children at baseline, suggesting lower overall vagal regulation of heart rate; (2) a large

percentage of children with autism had abnormally high sympathetic activity, i.e. skin conductance response. In another study [17] galvanic skin response data were compared between children with autism and typically developing ones. The study revealed that children with autism have more irregular patterns of skin conductance physiological signals.

Similarly, for a voice channel, you may consider collecting a sample from the child prior to the intervention or at the very beginning of the session to optimise the child's voice activity detector for the voice of the specific child of interest. For this, voice material of an approximate duration of 1 minute from a child would be ideal. Here are some suggestions on how to collect such material: (1) Older children might tell about their typical school day, their last holiday, or a dream they had recently (without being interrupted by others in the room); (2) Younger children who tolerate headphones could hear some imitation prompts through headphones, e.g.: "Please repeat the following sentences: 'My friend has an apple tree in his garden', 'Kaspar loves music and meeting new friends'"; (3) Younger children who do not tolerate headphones might hear these imitation prompts from a therapist or researcher who pauses the audio recording whenever the therapist herself/himself talks. To collect this voice material might even be embedded in a kind of a game or routine to "activate" the robot: "Please repeat all the sentences that I will tell you now to wake up Kaspar so that he can play with you/us funny games."

Of course collecting such voice material will not work for nonverbal/minimally verbal children and not for children who are shy/unhappy or have not acquired (yet) verbal imitation or short story-telling.

Guideline PROC6:
Combine methods to label the data with emotional states correctly.

As children with ASD often have different, partly unique, ways to express their emotions, it can be difficult to label their emotional states correctly. Therefore, a combination of diverse methods is recommended such as asking the child how he or she feels, expert-annotation of the data, asking caregivers who know the child well to evaluate the annotations, and labelling according to a stimuli tag.

Assignment of a proper (accurate) label to the data gathered is a challenge in emotion recognition in general, not only in autism. The internal emotional state (ground truth) of an individual is hard to determine in a continuous manner, even by himself/herself. Several strategies might be used for labelling the data. In the study, [18] skilled therapists annotated recordings from multiple therapy sessions, which was the most common practice. Another approach that uses subjective reports of the affective states from caregivers was introduced [19] and compared to therapists' reports with a consistency of approximately 83%. In a single study of human-robot interaction, loop [20] self-report was used, however, combined with therapists' evaluations. Self-report in children with autism was only partially consistent with tagging by therapists. Therapists' reports were taken as a "ground truth" for classification. The authors state that due to the deficits in

communication skills in children with autism, the "classic" methods for emotion tagging are hard to apply. They recommend that for enhancement of reliability of tagging, both a clinical observer and a caregiver who knows the participant shall be included in the study. [20]

When planning a study on emotion recognition during robot-child interaction, it is recommended to carefully reflect on the annotating strategy needed for the reliable labelling of the data to achieve the respective study purposes.

Guideline PROC7:
It might be hard to obtain general models as characteristics and deficits of children with autism differ.

Autism is a heterogeneous disorder with various symptoms and severity levels. For example, some children have good verbal skills, whereas others are nonverbal. Moreover, robot-supported intervention can be a valuable alternative to standard therapeutic approaches for children of different ages. Of course, activities for 3-year-old children should not be the same as for 12-year-old children due to differences in their interests and capacities. It would be necessary to implement emotion recognition technologies that are suitable or adaptable for different languages, cultural backgrounds, and the gender of the child. Optimally, such technologies should be robust to diverse acoustic environments, microphone types and positions, eye tracking devices, types of robots, etc. More studies are needed to evaluate how different settings and devices influence emotion recognition performance. A large amount of annotated data – recorded in different rooms, situations, languages, robots, etc., is helpful in optimising existing emotion recognition models.

4.6. Emotion recognition in children with autism (EMO)

Among the deficits observed in autism spectrum disorder, one might outline challenged expression and recognition of emotional states. With regard to automatic emotion recognition, atypical expressions of a child might result in false assumptions regarding its internal emotional state. Most of the guidelines in this section are informative only, as there is no way to change how symptoms of emotions occur.

Guideline EMO1:
Children with autism exhibit atypical expression of emotions

Children with autism show roughly the same level of intensity of emotional facial expression, but their patterns may be atypical. For example, significant differences were shown between high-functioning individuals with autism and typically developing individuals for disgust and sadness across the face, joy across the upper and lower parts, and surprise only across the lower part of the face. In contrast, no

significant differences were found for the emotions of anger and fear [11]. Therefore, generic emotion recognition methods trained on data for typically developing children might be only partially appropriate. It is recommended to train more specific classifiers on data gathered from children with autism only instead of general ones for all children.

Guideline EMO2:
Children with autism exhibit some atypical synchronisation of voice/facial/gestures expression of emotions

Several studies are reporting that children with autism exhibit atypical symptoms of emotions. A study of cross-modal coordination of emotion expressions reveals that the coordination is lower in the ASD group when compared to neurotypical children [21]. Children with ASD produce the same level of emotional facial expressions and speech at the same intensity levels as typically developing children, but facial and vocal expressions are less coordinated with each other. According to the literature review summarised in the same paper, that cross-modal coordination also applies to facial expressions versus gestures. It is also reported that children with ASD exhibit atypical timing and synchrony of movements of different facial regions, reduced intensity of upper face movements, reduced variety of facial movements, and more ambiguity, as expressions for positive and negative valence do not differ as they do for the typically developing peers. Another study confirmed the observations [22], which reported less synchrony of motions between facial expressions, less complex facial dynamics, and more ambiguity. The deficit in facial expression was independent of emotion type (happiness, anger, sadness, and neutral state were included).

Guideline EMO3:
Children with autism might exhibit no or little speech and vocalisations in the interaction with a robot

As known from the literature, most children with autism have - partly severe - difficulties in speech-language acquisition [14, 15]. Therefore, it is likely that at least some of them will exhibit little or no speech in interactions with social robots. This has been demonstrated recently by Milling et al. [10] who applied a deep learning based voice activity detector that detected child vocalisations in only around 4% of recorded child-robot-intervention sessions. In this study, the session time with detected child vocalisations ranged from 0.6% to 20%. Emotion recognition based on speech signals might be more successful for children with higher speech-language skills than children with lower speech-language skills.

Guideline EMO4:
Children with autism often look sideways or down when speaking

Several studies report children with autism exhibit atypical behaviours that might influence the recordings. That applies especially to interactive activities. For example, a child frequently looks sideways or down during a conversion. The action refers not only to an eye gaze but also to head-turning. This might influence capturing both facial expressions and areas of interest from the eye-gaze.

As the child's condition is given, and we have a limited ability to influence it, the recommendation is just to report such behaviours.

Guideline EMO5:
Have in mind that you are capturing symptoms of emotions only

The activation of the human body's nervous system induces changes in life activities,which might be interpreted as symptoms of emotions. However, out of the life activities and modalities used as proxy symptoms for emotion recognition, all are prone to some disturbance and misinterpretation.

The internal emotional state (ground truth) is hard to determine, sometimes even for someone who experiences it. There is no way of telling what is the current emotional state of a human being, and even self-reports might be biased.

However, it is important to consider that the "ground truth" on the internal phenomena of emotions remains unknown, and all labels are biased [12]. Some of the emotional expressions might be changed by a person purposefully. For example, children with autism often smile when left alone, not because they feel joy.

Therefore, interpretation of the outcomes (estimations provided by the algorithms) in emotion recognition should refer to symptoms rather than emotions attributed to them. It would be more accurate to refer to emotional expressions rather than to a person's emotional state, having in mind that we could observe and measure symptoms only. Automated solutions should keep the human in the loop.

4.7. Design of research studies (RES)

Apart from therapists, researchers are also among the intended audience of those guidelines. Therefore, this section refers specifically to research-based studies.

Guideline RES1:
Consider the construct of the study and the control group.

Creating a study and control group requires addressing several challenges. To begin with, it is difficult to balance the proportion of children participating in the study by gender due to the fact that autism is far more commonly diagnosed in boys than girls. A

similar problem may arise when balancing the group of children in terms of low and high functioning. Low-functioning children, depending on the type of activity, may be more likely to refuse to participate. When constructing a control group, children's intellectual abilities should be considered. Therefore, the control group should be formed based on developmental age and level of functioning rather than the chronological age of children in the group with ASD. [2]

Guideline RES2:
Create an openly available and well described dataset for future studies.

The affective computing discipline benefits from many datasets that are recorded and openly shared among researchers. The sets contain feature sets or row data, anonymized, followed by emotion-related labels or values to be used to classify the emotions. However, there is only a single dataset available for emotion recognition applied to children with autism. It was created by the deENIGMA project and might be used in future research [23]. It contains recordings from 35 children with autism.

While preparing and publishing a dataset, following the FAIR rules is advisable for open science [35]. FAIR is an acronym that stands for the following dataset features: Findable – Accessible – Interoperable - Reusable. In order to be operable, the data should use a formal, accessible, shared, and broadly applicable language for knowledge representation, use proper vocabularies for metadata and include references to other related resources. To be Reusable, metadata shall be richly described with a plurality of accurate and relevant attributes, be released with a clear and accessible data usage licence, and meet domain-relevant community standards. To be Findable, the dataset should be assigned with a globally unique and persistent identifier, and it should be registered or indexed in a searchable resource. To be Accessible the dataset must be retrievable by the identifier using a standardised communications protocol (with the protocol being open, free, and universally implementable).

We would like to emphasise that sharing the data openly requires obtaining written consent from participants or caregivers. Therefore, privacy and anonymity must be secured. Moreover, it is advisable to obtain an approval from an ethical committee (see GEN3).

Guideline RES3:
Report children and children groups characteristics and be as detailed as possible.

As autism deficits occur in various forms and levels in individuals, the challenge of generalizability of the results remains inherent in all of the studies. Therefore, regardless of the group construction, in each study, it would be valuable to provide information on participants as detailed as possible, including at least: chronological and developmental

age, gender, level of functioning and comorbidity, followed by therapy length and background.

Guideline RES4:
Report familiarisation sessions and sessions that failed.

Be prepared that data would be simply unavailable for some children and/or sessions. The reasons behind the failure might be diverse - starting with technical problems (such as disconnection of the sensor, device failure or low battery), through human errors (not turning the device on, calibration mistake) to the manifestation of a child's specific behaviour under observation. For example, a child might refuse to wear a wristband, or other wearable device, refuse to interact with a robot, exhibit some anxiety tantrum, or simply take a body posture that would hinder capturing the data (for example sit sideways).

All the cases of an interaction failing for this reason or other should be reported along with the cause. This would provide transparency to the study, making it more reliable and valuable for future research. In addition, it is very useful if researchers describe what did not work, as it allows them to identify blind paths, useless devices or techniques, and challenges one might encounter in a similar study.

Guideline RES5:
If evoking emotions in your study, choose appropriate stimuli

Some research studies might require evoking emotional states rather than observing natural ones. There were multiple strategies used in the studies to evoke emotions (stimuli) and tag them. Regarding the stimuli, both evoked emotions and a natural interaction approach were used, and the challenge of collecting a proper training set of data from children with autism was raised [24]. The most common stimuli used in the studies included pictures and video resources. However, a study [17] on galvanic skin response revealed that pictures are not suitable stimuli for evoking emotions in children with autism. In another study, eye gaze patterns were analysed as a reaction to stimuli of videos containing human faces. Although video stimulus was the most frequently used one, other studies tried: a serious game approach [25], computer-based intervention tools [26], or observation of natural human-robot interaction [20] [27].

4.8. Reporting studies on children with ASD (REP)

While reviewing papers, we have noticed some ambiguity regarding terms and phrases used. This section refers to reposting studies, mainly defining and using acceptable terms.

Guideline REP1:
Distinguish the two meanings of "emotion recognition" phrase.

In the context of emotion recognition technologies supporting therapy of children with autism, the "emotion recognition" phrase has two meanings: emotion recognition by children with autism and recognizing emotions of children with autism. The first one refers to the ability of children to recognize emotion in others. The latter concerns automatic emotion recognition technologies applied to recognize and analyse the internal emotional states of a child.

Guideline REP2:
Be precise to describe devices, channels and modalities, be aware of distinction between those.

When referring to inputs used for emotion recognition, it is important to distinguish between life activities, observation channels, and modalities.

The process of emotion recognition is analysed with respect to **life activities** i.e. conscious and unconscious actions of a human body, which generate a specified symptom of an emotional state, that can be further analysed in emotion recognition. The following life activities were analysed in the selected papers: various types of movement, a sound made by humans, physiological activities: heart activity, unconscious muscle activity, respiration, and thermal regulation. In addition, the activation of the human body's nervous system induces changes in life activities which might be interpreted as symptoms of emotions.

The life activities might be recorded via **observation channels**, which are mediums for recording a signal holding information on observable symptoms. The channel refers to a type of signal obtained rather than a physical medium. The channels that were used in the studies of emotion recognition in children with autism include RGB video, depth video (Kinect mainly), audio, ECG (electrocardiography), BVP (blood-volume pulse), chest size, EMG (electromyography), fMRI (functional magnetic resonance imaging), EDA (electrodermal activity), and temperature.

The life activities generate **modalities**, which are understood as information observable and used as a proxy for emotion recognition. In our study, modalities are grouped according to life activities:

- ➢ movement: facial expressions, body postures, eye gaze, head movement, gestures

(also called hand movements) and any other not previously classified motion;
- ➢ sound expressions: vocalisations, the prosody of speech;
- ➢ heart activity: heart rate, HRV (heart rate variability);
- ➢ muscles activity not related to movement: muscles tension;
- ➢ perspiration: skin conductance;
- ➢ respiration: intensity and period;
- ➢ thermal regulation: peripheral temperature;
- ➢ brain activity: neural activity. [2]

We recommend first defining observation channels and providing detailed information on devices used for recording them. Further, report modalities obtained via observation channels.

Guideline REP3:
Be considerate with the use of terms that refer to children.

Inclusive language acknowledges diversity and conveys respect to all people. Please pay attention to wording about the children so as not to imply that one individual is superior to another on the health condition or disorder. During our studies, we noted that some studies refer to individuals with autism as "autistic children", while to the individuals in the control group as "normal children", which should be avoided [1, 2]. We suggest using more appropriate alternatives such as "typically developing children" or "neurotypical" instead. When referring to children on the autism spectrum, it is recommended to put a person first, for example, "a child with autism" or "individuals on the autism spectrum".

Guideline REP4:
Name and define emotional states addressed

As psychology does not properly define emotions, researchers, teachers and trainers use diverse labels to name them. Some authors refer to emotional states from the basic set or other ones with their own labels, and those labels are not further defined from a psychological perspective. Some terms used by authors could be grouped. For example, happy, happiness, joy, and smile are all different, but they are used interchangeably in most studies. The guideline refers to naming the recognized state and defining it properly. For example, if you recognize a smile, please refer to it as a smile, not happiness or joy. Another example is the fear-related emotions group with fear, anxiety, and trepidation. None of those terms is explicitly defined in the studies.

Some states addressed in the studies [36-38] are more attention-related than emotion-related (engagement, involvement), and that seems more of the real interest in studying children with autism (and not six basic emotions). An interesting concept of compound emotions, such as fearful surprise or happily disgusted, was also raised in the context of the meltdown crisis of a child.

As psychologists define particular discrete emotions differently, it is recommended to choose one of those definitions for your study and report it.

5. Guidelines evaluation

Multiple methods of evaluation have been applied in order to evaluate the final product of the project - ER-RIA Guidelines for Emotion Recognition in Robot-supported Interventions in Autism, including:

- questionnaire to obtain quantitative data;
- focus groups to obtain qualitative data;
- expert evaluation with AGREE instrument - both qualitative and quantitative.

5.1. Questionnaire

The first version of the guidelines was evaluated and then improved (list of changes - see section 5.4). The questionnaire regarded each of the guidelines separately. Each guideline was evaluated in terms of the following criteria:

- Adequate amount of description with a 5-point symmetric scale ranging: too little - too much (with 3 being the best grade);
- Understandability of the guideline and its description - with a 5 point agree-disagree scale (with 5 being the best grade);
- Applicability of the guideline - with a 5 point agree-disagree scale (with 5 being the best grade).

The questionnaire was handed over to 49 participants, who were asked to read guideline by guideline and answer the three questions per guideline. The participants were students who joined training on affective loop in robot-child interaction in autism therapy.

Questionnaire results might be summarised as follows:

- for too little - too much description, the average score was: 3,25 +- 0,68 (1 is too little, 3 is neutral 5 is too much);
- too little description (with >=10 people rating 1 or 2) was pointed out for the following guidelines: GEN3, CH1, INT2, EMO4, RES3;
- too much description (mean score >3,5) was obtained by the following guidelines: GEN2, CH3, SYM1, SYM10, TECH5, INT1, PROC2, PROC6;
- average understandability for all guidelines was 4,57 +- 0,69 (5 - strongly agree);
- less understandable (<4,5): GEN1, GEN2, TECH3, INT1, PROC2, PROC5, PROC6, PROC7, EMO2, RES3, RES5, REP2;
- only one guideline was rated under 4 with regard to understandability: PROC2 (3,97+- 1,05);
- average applicability for all guidelines was 3,99 +-0,98 (5 point Likert scale, 5 - strongly agree, 3 - neutral)
- relatively lower applicability (average under 4) was scored for guidelines: GEN1, GEN3, CH1, CH4, TECH4, TECH5, INT1, PROC1, PROC2, PROC3, PROC4, PROC5, PROC6, PROC7, EMO1, EMO2, EMO3, EMO4, EMO5, RES2, RES5, REP4;
- very low applicability (under 3,5) was obtained for: PROC2, PROC6, EMO2 - all of those had relatively low understandability as well.

Having the questionnaire results, we focused on the guidelines that were scored significantly lower and improved them to obtain 1,1 version of the guidelines.

5.2. Focus group

After getting familiar with guidelines and handling questionnaires, participants were invited to join focus groups. Each group had 6 up to 8 participants, and there were 7 groups in total. Each of those groups had to answer the following questions:

- Identify 5 items (guidelines) that are the least understandable (with justification)
- Identify 5 items (guidelines) that are the hardest to apply (with justification)
- List 3 guidelines that are the most valuable for therapists
- List 3 guidelines that are the most valuable for developers
- List 3 guidelines that are the most valuable for researchers
- Could any of the guidelines be removed (is not necessary)?
- Could any guideline be added to the list?

Focus groups results might be summarised as follows (the numbers in parenthesis indicate the count of focus groups that mentioned the issue):

- large amount of guidelines (2), add page numbers (1);
- repetitions (3), contradictions (3)
- the least understandable: PROC2 (4 times), PROC5 (3 times), GEN3 (3 times), single occurrences: RES3, SYM10, TECH3, PROC7, SYM1, INT1, REP2, INT5
- the hardest to apply: all EMO guidelines (3), PROC2 (2), PROC5 (2), RES2 (2), REP4 (2), single occurrences: CH4, PROC3, PROC6, PROC7, TECH4, SYM3, SYM6, SYM9, REP3, RES3;
- none of the groups suggested removal of any guidelines, they rather suggested merging: CH1 + CH2 + CH3 + SYM2, SYM4 + SYM5, SYM7 + SYM8, merge SYM6 + SYM7 + SYM8, INT2 + TECH5, GEN3 + TECH1.

Focus groups provided a lot of useful qualitative information - not only on what to change, but how to improve the descriptions. Having the focus groups results, we improved guidelines according to the remarks (most of them) to obtain 1,1 version of the guidelines.

5.3. AGREE expert evaluation

Then the guidelines were evaluated by 3 experts using AGREE (The Appraisal of Guidelines for Research and Evaluation), which is an instrument to evaluate the process of practice guideline development and the quality of reporting. The AGREE II refined version was used - it comprises 23 items organized into 6 quality domains plus 2 general items [39]. The domains are: scope and purpose, stakeholder involvement, rigour of development, clarity of presentation, applicability, and editorial independence. The results might be summarised as follows:

- the general “quality of this guideline” item had a score of 6,12 (using 1-7 Likert’s scale);
- the general “recommend this guideline for use” item had a score of 2,9 (using 1-3 scale);
- out of 23 items 20 were scored over 6 (7- point scale);
- the following items were scored lower than 6:
    - a procedure for updating the guideline (4,95),
    - the guideline describes applicability (5,73),
    - the guideline presents evaluation criteria (5,02);
- qualitative remarks included (among others):
    - a statement that the funding body didn't influence the content of the guidelines should be added next to the funding acknowledgement;
    - adding some references to improve body of knowledge visibility in guidelines;
    - RES group of guidelines is more general and applies not only to robot-child studies in autism therapy;
    - EMO guidelines are more descriptive in nature and do not contain remedies or recommendations what to do.

Having the AGREE instrument results, we improved guidelines according to the remarks (most of them) to obtain 1,2 version of the guidelines.

5.4. Changes - guidelines 1.0 and 1.2

Having the guidelines evaluated with the three method, we have significantly improved the descriptions. Some major changes included:

- merging guidelines (49 in version 1.0, 46 in version 1.1 and 44 in version 1.2);
- addressing repetitions and contradictions between the guidelines; adding page numbers;
- adding a statement on independency of guidelines development from funding body;
- adding more information on how the guidelines was developed (section 3);
- adding a section on evaluation and monitoring of the guidelines (section 5);
- adding information on applicability (section 6);
- improvement of descriptions of almost all guidelines (with special focus on the ones mentioned in focused groups, and those that scored less in questionnaire and/or AGREE instrument results).

6. Applicability

The guidelines are developed for three target groups (see section 2.2): autism therapists, technology developers, and researchers. We are aware that addressing this diverse audience might cause some applicability confusion - guidelines valuable for one group, might be neglected by the other one. Therefore we asked focus groups to point out the guidelines valuable for each of the target subgroup.

The guidelines that are the most valuable for **therapists** are as follows:

- GEN2 and GEN 3 to start with;
- all guidelines in SYM section to get familiar how symptoms of emotions are captured;
- all guidelines in INT section that provides hints how to plan and conduct interaction;
- all guidelines in TECH section, if missing technical person support.

The guidelines that are the most valuable for **technology developers** are as follows:

- GEN1 to put the child first and GEN 3 to comply with ethical requirements, such as privacy;
- all guidelines in PROC section - how to process the data;
- all guidelines in EMO section that describes the specificity of emotional symptoms expressed by children with autism;
- TECH2 to TECH5 - technical requirements for technologies developed;
- CH1 and CH2, INT2.

The guidelines that are the most valuable for **researchers** are as follows:

- all guidelines in RES section - those are the guidelines how to deal with studies on robot-child with autism interaction;
- all guidelines in REP section that describe how to report the studies;
- selected guidelines in CH and EMO sections - to understand limitations of work with children with autism and to be able to evaluate data quality;
- PROC5 and PROC7.

Regarding applicability, we are aware of the fact that some of the guidelines are not so easy to apply and that is a result of some inherent challenges in the domain. Looking at questionnaire scores, guidelines in sections: GEN, PROC, EMO, RES and selected guidelines in TECH and CH sections were scored lower with regard to applicability criteria.

The GEN section is a general one by definition and those are rather general recommendations of a start point nature. Then the following challenges make it hard to apply some of the guidelines:

- emotion phenomena is difficult itself to precisely describe and detect, and the challenge is even bigger while using automatic recognition technologies for affect classification;
- work with children with autism is challenging - starting with their deficits, but also distress behaviours and tendencies;
- emotion symptoms expressed by children with autism are different than the ones expressed by their typically developing peers;
- not only therapy, but also research studies with children with autism must follow a child, which is good for a child, but not good for research due to missing or low quality data and/or limited repeatability.

As a result of the above mentioned challenges, as long as we have tried to improve applicability and understandability of the descriptions, some applicability issues cannot to be eliminated,

for example getting the ground truth for child's emotional state, having a totally silent room for research, or creating a publicly available dataset.

7. Future works

This document is the product of an international project EMBOA funded by European Union programme Erasmus Plus. This document is distributed free of chargé on CC-BY open licence. The document is available at EMBOA project website http://emboa.eu/ in English, Polish, Macedonian, German and Turkish. The document is free to re-distribute.

Although the project ends in 2022, we plan to perform further research on the topic, and perhaps extended the guidelines. Some of the ideas of further development - some of them were suggested during evaluation process or resulted from project team observations - but were not addressed so far due to being outside of the scope of te project:

- developing guidelines towards future automated interaction in order to close the affective loop mentioned in the guidelines;
- define future technologies, rather than providing guidelines how to deal with currently available ones;
- provide recommendations for the cooperation between the target groups - researchers, therapists, and technology developers;
- divide the guidelines to "during design" - "before session" - "within session" - "post-hoc" categories.

If you want to work with the guidelines, have suggestion for their improvement or extension, or maybe want to translate them to your national language, please do not hesitate to contact Agnieszka Landowska, nailie@pg.edu.pl.

Literature

1. Bartl-Pokorny, K. D., Uluer, P., Barkana, D. E., Baird, A., Kose, H., Zorcec, T., Robins, B., Schuller, B., Landowska, A., & Pykała, M. (2021). Robot-Based Intervention for Children With Autism Spectrum Disorder: A Systematic Literature Review. IEEE Access, 9, 165433-165450. https://doi.org/10.1109/access.2021.3132785

2. Landowska, A.; Karpus, A.; Zawadzka, T.; Robins, B.; Erol Barkana, D.; Kose, H.; Zorcec, T.; Cummins, N. Automatic Emotion Recognition in Children with Autism: A Systematic Literature Review. Sensors 2022, 22, 1649. https://doi.org/10.3390/s22041649

3. Karpus A., Landowska A., Miler J., Pykała M.: Systematic Literature Review – methods and hints, ETI Faculty Technical report, Gdansk University of Technology, 1/2020,

4. Bartl-Pokorny K.D., Pykała M., Erol Barkana D., Baird A., Köse H., Zorcec T., Robins B., Schuller B.W., Landowska A. Systematic Literature Review - Robot-Based Intervention for Children with Autism Spectrum Disorder, ETI Faculty Technical report, Gdansk University of Technology, 2/2020

5. Landowska A, Wróbel M., EMBOA project evaluation report, ETI Faculty Technical report, Gdansk University of Technology, X/2022

6. Abdullah, S.M.S.A., Ameen, S.Y.A., Sadeeq, M.A. and Zeebaree, S., 2021. Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*, *2*(02), pp.52-58.

7. Landowska, A., Zawadzka, T., & Zawadzki, M. (2021). Mining inconsistent emotion recognition results with the multidimensional model. *IEEE Access*.

8. Yang, K., Wang, C., Sarsenbayeva, Z., Tag, B., Dingler, T., Wadley, G., & Goncalves, J. (2021). Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets. *The Visual Computer*, *37*(6), 1447-1466.

9. Milling, M., Baird, A., Bartl-Pokorny, K. D., Liu, S., Alcorn A. M., Shen, J., Tavassoli, T., Ainger, E., Pellicano E., Pantic, M., Cummins, N., Schuller, B. W. (2022). Evaluating the Impact of Voice Activity Detection on Speech Emotion Recognition for Autistic Children. Frontiers in Computer Science, 4, 837269.

10. Milling, M., Bartl-Pokorny, K. D., Schuller, B. W. (2022). Investigating Automatic Speech Emotion Recognition for Children with Autism Spectrum Disorder in Interactive Intervention Sessions with the Social Robot Kaspar. medRxiv. https://doi.org/10.1101/2022.02.24.22271443

11. Schuller, B. (2018). What affective computing reveals about autistic Children's facial expressions of joy or fear. *Computer*, *51*(06), 7-8.

12. Landowska A.: (2019) Uncertainty in Emotion recognition, Journal of Information, Communication and Ethics in Society, Vol. 17 No. 3, pp. 273-291, Emerald Publishing, DOI 10.1108/JICES-03-2019-0034

13. [Landowska, Miler, 2016] Landowska, A. and Miler, J. (2016), "Limitations of emotion recognition in software user experience evaluation context", Federated Conference on Computer Science and Information Systems, (FedCSIS), IEEE, pp. 1631 - 1640.

14. Belteki Z, Lumbreras R, Fico K, Haman E, Junge C. The Vocabulary of Infants with an Elevated Likelihood and Diagnosis of Autism Spectrum Disorder: A Systematic Review and Meta-Analysis of Infant Language Studies Using the CDI and MSEL. Int J Environ Res Public Health. 2022 Jan 27;19(3):1469. doi: 10.3390/ijerph19031469.

15. MacFarlane H, Salem AC, Chen L, Asgari M, Fombonne E. Combining voice and language features improves automated autism detection. Autism Res. 2022 Apr 23. doi: 10.1002/aur.2733.

16. Bal, E.; Harden, E.; Lamb, D.; Van Hecke, A.V.; Denver, J.W.; Porges, S.W. Emotion Recognition in Children with Autism Spectrum Disorders: Relations to Eye Gaze and Autonomic State. J. Autism Dev. Disord. 2010, 40, 358–370.

17. Fadhil, T.Z.; Mandeel, A.R. Live Monitoring System for Recognizing Varied Emotions of Autistic Children. In Proceedings of the 2018 International Conference on Advanced Science and Engineering (ICOASE), Duhok, Iraq, 9–11 October 2018; pp. 151–155.

18. Marinoiu, E.; Zanfir, M.; Olaru, V.; Sminchisescu, C. 3D Human Sensing, Action and Emotion Recognition in Robot Assisted Therapy of Children with Autism. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2158–2167.

19. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Affect Recognition in Robot Assisted Rehabilitation of Children with Autism Spectrum Disorder. In Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Rome, Italy, 10–14 April 2007; pp. 1755–1760.

20. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Online Affect Detection and Robot Behavior Adaptation for Intervention of Children with Autism. IEEE Trans. Robot. 2008, 24, 883–896.

21. Sorensen, T.; Zane, E.; Feng, T.; Narayanan, S.; Grossman, R. Cross-Modal Coordination of Face-Directed Gaze and Emotional Speech Production in School-Aged Children and Adolescents with ASD. Sci. Rep. 2019, 9, 18301.

22. Grossard, C.; Dapogny, A.; Cohen, D.; Bernheim, S.; Juillet, E.; Hamel, F.; Hun, S.; Bourgeois, J.; Pellerin, H.; Serret, S.; et al. Children with autism spectrum disorder produce more ambiguous and less socially meaningful facial expressions: An experimental study using random forest classifiers. Mol. Autism 2020, 11, 5.

23. Rudovic, O.; Lee, J.; Dai, M.; Schuller, B.; Picard, R.W. Personalized machine learning for robot perception of affect and engagement in autism therapy. Sci. Robot. 2018, 3, eaao6760.

24. Tang, T.Y. Helping Neuro-Typical Individuals to "Read" the Emotion of Children with Autism Spectrum Disorder: An Internet-of-Things Approach. In Proceedings of the 15th International Conference on Interaction Design and Children, IDC'16, Manchester, UK, 21–24 June 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 666–671.

25. Di Palma, S.; Tonacci, A.; Narzisi, A.; Domenici, C.; Pioggia, G.; Muratori, F.; Billeci, L. Monitoring of autonomic response to sociocognitive tasks during treatment in children with Autism Spectrum Disorders by wearable technologies: A feasibility study. Comput. Biol. Med. 2017, 85, 143–152.

26. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. Int. J.-Hum.-Comput. Stud. 2008, 66, 662–677.

27. Silva, V.; Soares, F.; Esteves, J. Mirroring and recognizing emotions through facial expressions for a RoboKind platform. In Proceedings of the 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG), Coimbra, Portugal, 16–18 February 2017; pp. 1–4.

28. Adolphs R, Mlodinow L, Barrett LF. What is an emotion? Curr Biol. 2019 Oct 21;29(20):R1060-R1064. doi: 10.1016/j.cub.2019.09.008. PMID: 31639344; PMCID: PMC7749626.

29. H. Gunes, B. Schuller, M. Pantic and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey," 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2011, pp. 827-834, doi: 10.1109/FG.2011.5771357.

30. Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions?. *Psychological review*, *97*(3), 315.

31. Rudovic, O.; Lee, J.; Dai, M.; Schuller, B.; Picard, R.W. Personalized machine learning for robot perception of affect and engagement in autism therapy. Sci. Robot. 2018, 3, eaao6760.

32. Gay, V.; Leijdekkers, P.; Wong, F. Using sensors and facial expression recognition to personalize emotion learning for autistic children. Stud. Health Technol. Inform. 2013, 189, 71–76.

33. Hirokawa, M.; Funahashi, A.; Itoh, Y.; Suzuki, K. Design of affective robot-assisted activity for children with autism spectrum disorders. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; pp. 365–370.

34. Javed, H.; Jeon, M.; Park, C.H. Adaptive Framework for Emotional Engagement in Child-Robot Interactions for Autism Interventions. In Proceedings of the 2018 15th International Conference on Ubiquitous Robots (UR), Honolulu, HI, USA, 26–30 June 2018; pp. 396–400.

35. Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

36. Akinloye, F.O.; Obe, O.; Boyinbode, O. Development of an affective-based e-healthcare system for autistic children. Sci. Afr. 2020, 9, e00514

37. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. Int. J.-Hum.-Comput. Stud. 2008, 66, 662–677.

38. Krupa, N.; Anantharam, K.; Sanker, M.; Datta, S.; Sagar, J.V. Recognition of emotions in autistic children using physiological signals. Health Technol. 2016, 6, 137–147

39. AGREE II. Brouwers MC, Kho ME, Browman GP, Burgers J, Cluzeau F, Feder G, Fervers B, Graham, ID, Grimshaw J, Hanna S, Littlejohns P, Makarski J, Zitzelsberger L on behalf of the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. Can Med Assoc J. Dec 2010, 182: E839-842; doi:10.1503/cmaj.090449