



Guidelines for Emotion Recognition in Robot-Supported Interventions in Autism

Michał R. Wrobel¹ · Katrin D. Bartl-Pokorny^{2,3} · Tatjana Zorcec⁴ · Duygun Erol Barkana⁵ · Hatice Kose⁶ · Manuel Milling^{3,7,8} · Ben Robins⁹ · Björn W. Schuller^{3,7,8,10} · Agnieszka Landowska¹

Received: 26 April 2025 / Revised: 29 April 2026 / Accepted: 6 May 2026
© The Author(s) 2026

Abstract

Therapy for children with autism spectrum disorder benefits from the use of diverse technologies, as most children with the condition are eager to interact with them. The integration of emotion recognition technologies with social robots offers promising opportunities to make therapeutic interactions more adaptive, personalized, and responsive to a child's emotional state, potentially improving engagement and learning outcomes. The purpose of this study was to evaluate the feasibility of combining robot-based interventions with emotion recognition technologies for therapeutic use in autism. Specifically, we sought to identify best practices and obstacles in the integration of these technologies. To this end, we developed a comprehensive set of guidelines, which includes recommendations on observation channels, reliable acquisition of emotional expressions, technologies and devices, data processing, study design and reporting, and the specificity of emotional expressions in individuals with autism. They were evaluated using a triad of methods: questionnaires, focus groups, and a formal expert review using the AGREE II instrument. The results indicate that the guidelines are understandable and adequately described, and the focus groups helped identify the most relevant recommendations for researchers, therapists, and technology developers. The guidelines presented in this paper constitute the final version, which was refined based on the evaluation of preceding iterations.

Keywords Autism · Robot-enhanced therapy · Social robots · Automatic emotion recognition

✉ Michał R. Wrobel
michal.wrobel@pg.edu.pl

Katrin D. Bartl-Pokorny
katrin.bartl-pokorny@medunigraz.at

Tatjana Zorcec
tzorcec@gmail.com

Duygun Erol Barkana
duygunerol@yeditepe.edu.tr

Hatice Kose
hatice.kose@itu.edu.tr

Manuel Milling
manuel.milling@tum.de

Ben Robins
b.robins@herts.ac.uk

Björn W. Schuller
schuller@tum.de

Agnieszka Landowska
nailic@pg.edu.pl

² Division of Neonatology, Medical University of Graz, Auenbruggerplatz 38, Graz 8036, Austria

³ CHI – Chair of Health Informatics, TUM University Hospital, Neherstraße 1, 81675 Munich, Germany

⁴ University Children's Hospital, Medical Faculty, University of Skopje, 50 Divizija no. 6, Skopje 1000, North Macedonia

⁵ Faculty of Engineering and Natural Sciences, Department of Electrical and Electronics Engineering, Yeditepe University, 26 Agustos Yerleşimi, Istanbul 34755, Türkiye

⁶ Faculty of Computer and Informatics, Department of AI and Data Engineering, Istanbul Technical University, Ayazaga Campus, Maslak, Sariyer, Istanbul 34469, Turkey

⁷ MCML – Munich Center for Machine Learning, Oettingenstraße 67, 80538 Munich, Germany

⁸ MDSI – Munich Data Science Institute, Walther-Von-Dyck Str. 10, Garching bei München, 85748 Munich, Germany

⁹ School of Computer Science, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, UK

¹⁰ GLAM – The Group on Language, Audio, & Music, Imperial College London, 180 Queen's Gate, South Kensington, London SW7 2AZ, UK

¹ Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, ul. Narutowicza 11/12, Gdansk 80-233, Poland

1 Introduction

Autism spectrum disorder is a neurodevelopmental condition affecting how an individual perceives the world and interacts with others. Characterised by deficits in social communication and the presence of restricted/repetitive behaviours, it affects approximately 1 in 36 children aged 8 years in the US [1]; worldwide, around 1% of children are diagnosed with autism [2]. Autism introduces challenges to independent living, self-care, education, and employment opportunities. Around 33% of individuals with autism have intellectual disability [2] and 25–30% stay minimally verbal or never develop speech at all [3, 4].

Although it is a lifelong neurodevelopmental condition, a range of therapeutic interventions can support learning and development [5]. It is crucial for interventions to prioritise the enhancement of a child's social skills, given the research findings that social competence serves as a predictor for the long-term outcomes of individuals with autism [6]. Autism is a highly heterogeneous disorder [7] requiring personalised and tailored interventions for each individual. What works for one child may not work for another [8].

Without appropriate intervention, autism may lead to family breakdown, mental health problems and a lifelong caring role for family members. Therefore, addressing the unique needs of individuals with autism through targeted interventions is critical to mitigating potential challenges and promoting positive outcomes.

Children diagnosed with autism struggle with a range of deficits, including limited social, emotional, and imitation skills, which significantly affect their engagement in communication and interaction [7, 9, 10]. These communication deficits affect their relationships with family members and peers and are also present in the interaction with therapists during interventions. However, contrary to interactions with humans, many children with autism seem to enjoy interactions with robots, even if a robot has a human-like appearance. The reason for the preference of robots is unclear, but it might be related to the fact that robots provide a consistent, predictable, and reliable environment [11]. Regardless of the underlying cause, social robots have proven effective in overcoming the social challenges faced by children with autism, enabling them to actively participate in interactions. Once the initial interaction has taken place, there is a clear opportunity to progressively engage the child in developing and practising social and emotional skills [12].

Improving social interaction with the assistance of robots could involve incorporating an emotional dimension and introducing an affective loop into the process. To establish this, it is essential not only to influence the emotions of children with autism but also to perceive and respond to their emotional cues accurately.

It is important to acknowledge that while robot-supported interventions show promise, they are not meant to replace human interaction or traditional therapeutic methods. Instead, they are intended to complement existing interventions and serve as additional tools to enhance learning, interaction, and development.

As research in this area continues to advance, we can expect even more sophisticated and effective robot-based interventions to be developed, further improving the support and outcomes for children with autism [13]. However, the involvement of trained professionals, such as therapists and educators, remains crucial in guiding and implementing these interventions to ensure their appropriate and effective use.

The use of robots in this context offers several advantages, including:

1. **Engagement and interaction:** Children with autism often find it challenging to engage in social interactions. Robots can provide a predictable and non-threatening presence, making initiating, and sustaining interactions easier for these children [14].
2. **Personalisation:** Robots can be programmed to cater to each child's specific needs and preferences, providing personalised learning experiences tailored to their unique developmental goals [15].
3. **Consistency and repetition:** Children with autism benefit from consistent and repetitive learning experiences. Robots can repetitively deliver activities and exercises without becoming exhausted or impatient, promoting skill development through consistent practice [16].
4. **Reduced anxiety:** Robots can create a controlled and less overwhelming environment for children with autism, reducing anxiety and sensory overload often associated with social interactions.
5. **Data collection and analysis:** Many robot-based interventions come with data collection capabilities that allow therapists and educators to track a child's progress objectively. This data-driven approach helps evaluate interventions' effectiveness and adjust strategies as needed [17].
6. **Generalisation of skills:** Robots can help children with autism generalise their learnt skills to real-life interactions. As children become more comfortable and proficient in interacting with robots, they may transfer these skills to their interactions with people in their daily lives.

Even though these advantages have been explored for a couple of years, the majority of studies have been exploratory and have methodological limitations [18]. Although the design of the robot has received considerable attention, other aspects such as interaction protocol, therapeutic goal

achievement, observational measures, or research validity have received very little scrutiny. And this is the research gap we want to address.

The goal of the study was to confirm the feasibility of using robot-based interventions and emotion recognition technology to treat autism. Additionally, we aimed to identify the best practices and obstacles associated with using these technologies together. As a result, we developed comprehensive guidelines that bridge the gap between the technological potential of robots and emotion recognition and their clinical applications and studies. To the authors' knowledge, no such set of guidelines has been proposed for these technologies. Most studies offer only fragmented views of specific technologies, and some emphasize the need for a more comprehensive and general approach.

The proposed guidelines were subsequently evaluated using a combination of techniques: questionnaires, focus groups, and a formal expert evaluation with the AGREE II instrument [19].

This paper details the development and evaluation of the guidelines. The evaluation was based on three criteria: adequate description, understandability, and applicability. The thesis of the paper can be summarized as follows: **The proposed guidelines are understandable, adequately described, and applicable.** We also identified application areas for various stakeholders.

The remainder of the paper is organised as follows: The related work section presents studies that influenced and inspired our study. Section 3 introduces the guidelines and how they were developed. Section 4 provides a research methodology description with regard to the evaluation process, starting with the definition of the criteria we use for evaluation and details of each of the three research methods used. Section 5 provides the evaluation process results, followed by a discussion in Sect. 6, and conclusions. Importantly, the guidelines themselves are included in Appendix A.

2 Related Works

Robot-supported interventions for children with autism have gained significant attention in recent years [20, 21]. Researchers have been exploring using robots as therapeutic tools to enhance social, communication, and cognitive skills in children on the autism spectrum [12, 22]. Several studies have focused on the effectiveness of robot-supported interventions for individuals with autism [23].

The social robot Kaspar, developed by the University of Hertfordshire, has been used in a long-term case study evaluation with 170 children [24]. Results of the case study examples show, among others, how the robot could

encourage the use of language, mediate child-child or child-adult interaction, complement the work in the classroom, and provide the opportunity for basic embodied and cognitive learning [24].

Robins and colleagues performed a longitudinal study with four children with autism who interacted with the humanoid robot Robota for several months. The quantitative analysis of the video material documenting the child-robot interactions showed an increase in the duration of pre-defined behaviours towards the later trials. The qualitative analysis revealed that children made certain imitation, turn-taking, and role-switch behaviours towards the robot and other people in the room [25].

Some studies have explored using robots to help children with autism recognise and understand emotions. Pop et al. [26] used the social robot Probo to train three children with autism to identify situation-based emotions. The children's performance improved with moderate to large effect sizes in identifying sadness and happiness [26]. In a study by Marino and colleagues, children with autism who received robot-supported intervention substantially improved contextualised emotion recognition, comprehension and emotional perspective-taking [27]. Recently, Holeva and colleagues reported that robot-supported intervention improved the psychosocial skills of children with autism as evaluated by neuropsychological testing and parent reporting [28]. A study by Erol Barkana et al. explored the methodology of multi-modal data acquisition for observing emotions in children with autism during interactions with the social robot Kaspar, involving participants from multiple countries. The authors highlighted key challenges related to data availability, task and environment setup, and individual variability, emphasizing the need for personalized emotion models and adaptive emotion recognition technologies [29].

In their systematic review, Sani-Bozkurt and Bozkus-Genc revealed that although children with autism mostly show positive reactions to robots as social partners, there is still a controversy related to the effectiveness of robots for the development of joint attention skills in children with autism [30]. Salimi and colleagues evaluated 19 randomised control trials (RCTs) on robot-supported interventions for people with autism and found that about two-thirds of them reported positive results regarding the effectiveness of this intervention approach [13]. In their meta-analysis of 12 RCTs, Kouroupa and colleagues found that robot-supported interventions for children with autism significantly improved social functioning, whereas it did not improve emotional or motor functioning [16].

This overview showed that although many studies demonstrate the potential benefits of robot-supported interventions for children with autism, it is important to note that the field is still evolving, and more research is needed to

explore the long-term effects, individual differences, and optimal implementation strategies of robot-assisted therapies for autism.

An approach to enhance robot-supported interventions for children with autism is to embed automatic emotion recognition. In our recent review on automatic emotion recognition in children with autism, we identified that most of the 50 studies focused either on selected emotions of Ekman's six basic emotions or on the valence-arousal model [31]. The most frequently analysed modalities were facial expressions, skin conductance, heart rate and heart rate variability, peripheral temperature, and speech. The observation channels used were video, audio, electrodermal activity, and electrocardiography. The most frequently applied AI approaches were SVM classifier, neural networks, and k-nearest neighbors. Most studies used unimodal approaches; early fusion at the feature level was the most frequently applied for multimodal recognition.

Robot-supported intervention studies face a number of challenges. These challenges are related to the characteristics of the target study group, characteristics of the applied social robots or observation devices, interaction process between a child and a robot, and further methodological issues [12, 29, 31]. Among the major challenges related to characteristics of the study group are the limited expressive language capacities [32, 33] and the limited attention span [34] of many children with autism. The heterogeneity of autism leads to difficulties regarding intervention design as one task may be appropriate for some children but too complex for others [35–37]. This is especially problematic if children are not assigned to subgroups based on their level of functioning. Pop and colleagues [26] and Kim and colleagues [34] expressed concerns that some children may not understand the animations of the robot. Furthermore, interaction with a robot may be difficult for some children; Yun et al. mentioned that children may need a therapist to help with the interaction [38]. Some reported challenges in studies with children with autism to refer to the appearance of the robot, such as the presence of distractors on the robot [39] or the robot's gender [40]. Wang et al. reported that some participants did not like to wear physiological sensors that were intended for data acquisition [41]. Studies also raised methodological issues such as challenges regarding data analysis in unstructured and unconstrained interaction settings [36], difficulties for researchers and even qualified therapists to determine the emotions of children with autism [31, 33], potential cultural influences on emotion annotations [42], temporarily unavailable data from certain observation channels due to the nature of a specific task, children's characteristics or technical issues [31], as well as the integration of data from different sensors [43].

3 Guidelines for Emotion Recognition in Robot-Supported Interventions in Autism

Automatic emotion recognition technologies have received increasing attention over the last two decades and have enabled more and more application contexts [44]. The most common modalities applied in this context are videos of facial expressions, recorded speech, and physiological signals, all of which are prone to different types of noise and/or non-reliable availability [45, 46]. The level of these disturbances and the way emotions are expressed also depend on the individuals, their cultural background, and the context [47]. In particular, for children with autism, the use of emotion recognition is still in the exploration phase, and reliable and robust solutions are not developed yet [31]. Within this context, using robots with emotion recognition technologies has recently been introduced to the domain and is a novel approach designed to support and stimulate children with autism in strengthening their social skills [48].

To advance the use of automatic emotion recognition technologies in robot-supported intervention in autism, this paper introduces a comprehensive set of guidelines including recommendations and challenges in the field. We work towards achieving an affective loop in robot-based interventions, i. e., a combination of perceiving and stimulating the emotions of children with autism.

In the envisioned scenario, the robot would be able to react in real-time to the child's emotional state, which would allow a natural interaction with the child, preparing him or her better for real-life conversations with 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, allowing the therapist to concentrate more on the child to maximise the intervention session's success. For this purpose, an accurate and robust automatic emotion recognition solution is the key to supporting the children, caregivers, and therapists during these robot-assisted intervention sessions.

3.1 Target Groups

The guidelines cater to a heterogeneous audience, encompassing therapists, caregivers, researchers, and technology providers. The guidelines can support the therapists and caregivers with the possibility of enhancing therapy for children with autism with new technologies such as social robots, potentially combined with automatic emotion recognition. The researchers and experts from different backgrounds, including technological and pedagogical domains, may benefit from our observations and recommendations based on previous studies to organise, set up, conduct and report their own studies. Likewise, technology providers

might find alternating use cases on demand for enhancing social robots with automatic emotion recognition technologies. Furthermore, emotion recognition technology providers might benefit from these guidelines by identifying the potential challenges and limitations of using the technologies to perceive emotions in children with autism.

3.2 Development Methodology for the Guidelines

The development of the guidelines involved several steps, as shown in Fig. 1. First, two systematic literature reviews (SLR) were conducted to support the development of the guidelines. The first SLR explored the state-of-the-art combination of autism therapy and automatic emotion recognition technologies [31]; the second SLR focused on robot-supported intervention for children with autism [12]. In the next step, a first round of observational studies of robot-assisted intervention was completed. Based on our insights from the first two steps, we proposed a set of guidelines, labeled as version 1.0, and categorized them during brainstorming sessions. These guidelines were then evaluated using questionnaires and focus groups, as well as during the second round of observational studies. This allowed us to formulate version 1.1 of the guidelines. This version of the guidelines was evaluated using the AGREE II instrument and the final version 1.2 was prepared.

3.3 Observational Studies

The development of the guidelines benefits from two rounds of observational studies in which emotion observation technologies were used during robot-supported interventions. These practical observations were integrated with findings from the systematic literature reviews to propose the first version of the guidelines. The studies were held within the EMBOA project in Macedonia, Poland, the United Kingdom, and Turkey [29]. The observational studies were conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Gdansk University of Technology. Common interaction scenarios and protocols were used to conduct these observation sessions to ensure consistent studies across centers.

During the observational studies, the social robot Kaspar was utilized, engaging directly with the child participants. For these interactions, robot was operated under a Wizard-of-Oz paradigm, allowing researchers to control its responses in real-time. It is important to note that emotion recognition capabilities were not integrated into the robot's control system during these specific studies. Instead, the emotion recognition framework was developed in parallel to the observational studies, for consideration in future integration and autonomous operation of the robot [49, 50].

General assumptions for observational sessions were decided as follows:

- all setups should be ready before a child enters the room;
- multiple sessions might be held with a child
- familiarisation sessions are encouraged;
- there is a prepared common list of interaction scenarios to perform;
- the scenarios are translated into national languages and adjusted, if necessary (for example, children sing different songs in different countries);
- the specified scenarios are followed, and other interactions in between are allowed to be added on the run if/when needed;
- the video (with 2 cameras, if possible), eye-gaze, voice (with 2 microphones, lapel and general one), physiological signals (heart rate and skin conductance) are recorded;
- the data (using a randomly assigned alphanumeric code for each child) was recorded, stored, and shared within the consortium.

The list of prepared interaction scenarios is presented in Table 1. Each scenario contained basic prompts and reinforcements: “well done”, “bravo”, “thank you”, trumpet sound+“bravo”, “it hurts”, “your turn”. Those are used to encourage a child to follow the scenario and interact with Kaspar. The session started with “Hello” and ended with “Bye”. The order of the scenarios might be adjusted according to the respective intervention situation.

All involved researchers in the four countries performed the studies using the same equipment for conformance with observations. The equipment used in our observation studies is given as follows:

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

The particular devices were selected at the beginning of the observation studies after consideration of the literature review. Please note that the guidelines do not suggest any of those devices, but they were provided for future study applications (and perhaps replication). Here are the criteria considered for device selection:

- the lowest intrusiveness of measurement for a child;
- a possibility of long-term measurements;
- robustness to disturbance;

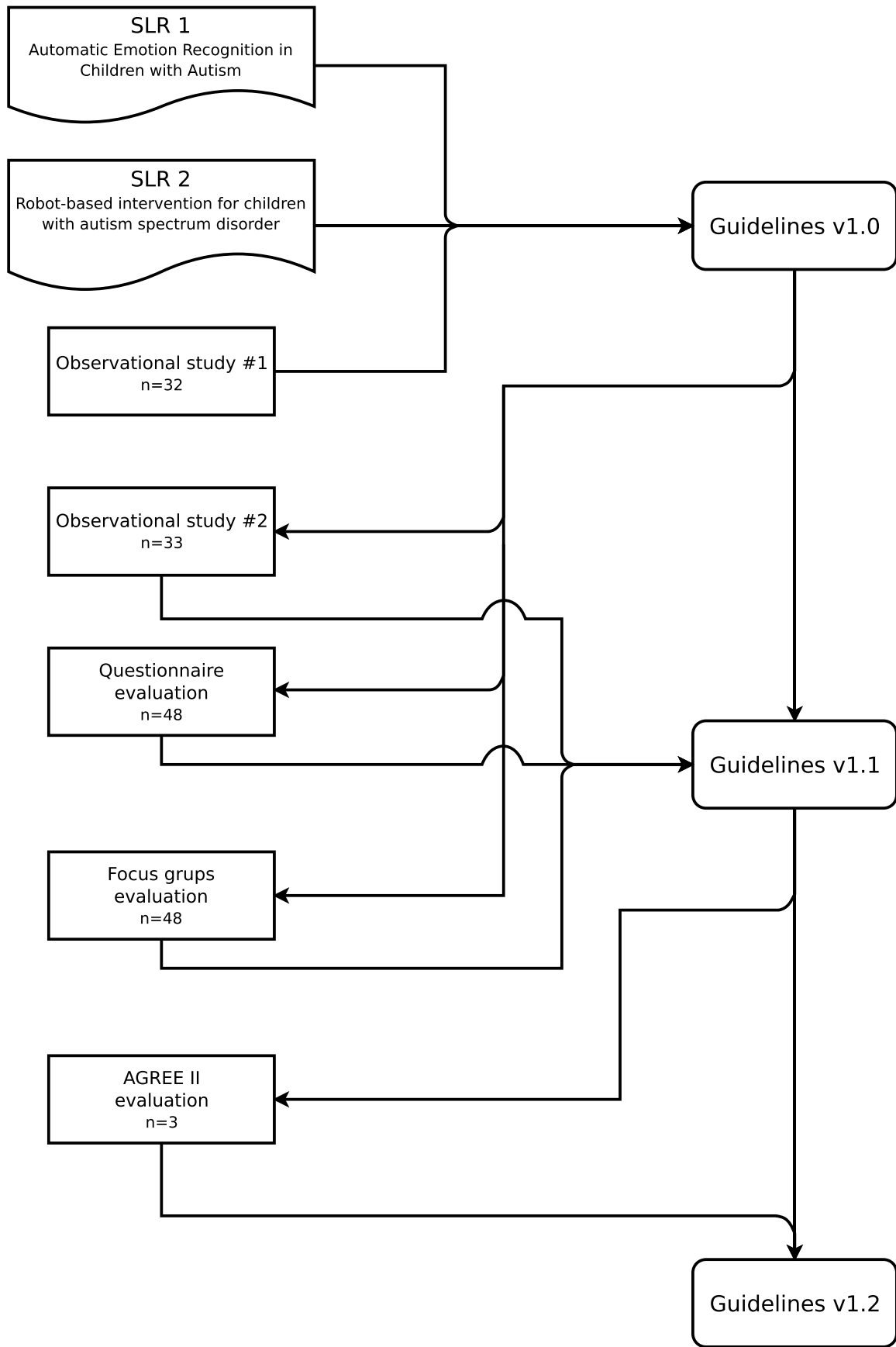


Fig. 1 Guidelines development process

Table 1 Interaction scenarios used in the observational study with the social robot Kaspar

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

Table 2 Observational study participants

Country	# of Children (Gender)	Age Range	# of Sessions (Min,Max)
Great Britain	22 (20 M, 2 F)	6–12	(2,3)
North Macedonia	15 (13 M, 2 F)	2–6	(2,11)
Poland	8 (8 M)	4–7	(1,4)
Turkey	20 (18 M, 2 F)	6–10	(1,2)

- data export function;
- quality to price ratio.

65 children with autism, aged 2 to 12, participated in the study, whose demographic data are shown in Table 2. Although the study protocol aimed for every participant to engage in all six scenarios, the interaction flow was adapted to individual needs. Due to the heterogeneity of the autism spectrum, including differences in developmental age and current emotional state, strict adherence to all scenarios was not always possible. Consequently, some sessions were shortened or specific scenarios were skipped in instances where a child exhibited distress, anger, or shyness, prioritizing their well-being over the completion of the full protocol. The following information was collected about the participants: developmental age, type, length of therapy, comorbidity, having siblings, rating of language understanding, verbal skills, animal knowledge, and body parts knowledge.

Given the young age of the participants and the sensitive nature of the collected information, particular attention was paid to ethical and legal considerations regarding data protection. Children are recognized as a vulnerable group and their data is protected under GDPR and EU AI Act [51–53]. The AI systems affecting children are subject to extra legal obligations. In EMBOA project, the sensitive multimodal affective data from children with autism is used. For extra protection of this sensitive data, the data is stored in local and highly protected servers and shared only with the responsible project workers under extra caution. Parental consents were taken for all participants, and the experimental

studies were ethically investigated and approved. The data was anonymized or pseudonymized for identity protection as well. Other techniques such as encryption or masking can be used for extra protection. Using the sensitive data of children is highly risky according to AI Act, and tasks such as emotion recognition will be avoided unless scientifically justified and ethically approved, as in the EMBOA project. Future works should also take these regulations and laws into consideration and limit their actions accordingly, such as minimal data production, and sharing with extra precautions, local storage and processing.

3.4 Brainstorming Sessions

During and after the observational study, the experimenters noted all remarks, notices, and lessons learnt as loose notes shared among the consortium. The sources of observations were diverse – we noted practical clues on using equipment, our observations regarding child-robot interaction, and remarks from therapists, parents, and technical operators. We also analysed the training sessions post-hoc, finding flaws and looking for potential solutions. After observational studies performed in North Macedonia, Poland, Turkey, and the UK, we had a large set of notes, not organised in any specific way. We decided to perform a brainstorming session to organise and structure the obtained content.

The brainstorming session involved six members of the EMBOA consortium from diverse international backgrounds. This multi-disciplinary group comprised an occupational therapist, affective computing researchers, robotics specialists, clinical linguist with experience in behavioural research, and experts in robot-assisted autism therapy. All participants have been previously involved in the literature reviews and observational studies. The session was conducted online, using the Lucidspark canvas tool and JITS meeting. The pre-session preparation required each participant to recall the project's purpose, scope, and target groups, and to review the results of the systematic literature reviews.

Additionally, participants were asked to independently review and revise their own individual notes collected during the observational studies. Then, the brainstorming session was divided into three phases:

- Phase 1. Every participant added proposed guideline statements as notes on the canvas. (no reading others' notes, silent phase);
- Phase 2. Reading others' notes. Giving thumbs up, adjusting own entries, making comments, moving notes together if they fit together (silent phase);
- Phase 3. Adding notes to categories, discussing the notes, merging (discussion phase).

The completion criteria for the 1st phase was that anyone has nothing more to add. For the second phase, each participant read all of the notes added by others and had nothing more to add. For the third phase, completion criteria were: all notes were assigned to areas, similar notes were grouped, and participants had nothing more to add. None of the individual notes was deleted during the session, but grouped, commented on, or re-phrased if unclear.

Post-session work included transforming the canvas into a textual representation of guidelines, adding a description for each guideline, and reviewing the compilation. Additional work was adding introductory and summary information to the guidelines document (purpose, scope, target audience, development description, etc.).

3.5 Description of Guidelines

The recommendations from the literature reviews and the lessons learnt from the observational sessions were integrated and summarised as guidelines. They have been divided into the following categories to systematise them and make them more accessible:

1. **General guidelines (GEN):** This section outlines fundamental principles and overarching considerations that should guide all aspects of emotion recognition in robot-supported interventions for children with autism. These guidelines prioritize ethical considerations, align technological applications with therapeutic goals, and establish a foundational understanding of emotional states.
2. **The choice of observational channels (CH):** This section focuses on selecting appropriate sensory inputs (e.g., video, audio, physiological) for capturing emotional expressions, considering factors like the intervention context, child's condition, and the trade-offs between data richness and intrusiveness. The aim is to optimize the collection of reliable emotional cues while minimizing disruption.
3. **Reliable acquisition of emotional symptoms (SYM):** These guidelines detail best practices for ensuring high-quality, consistent recording of emotional symptoms, such as facial expressions and vocalizations, by addressing environmental factors, sensor placement, and managing potential interferences. They aim to maximize the clarity and completeness of recorded data for accurate analysis.
4. **Technologies and devices (TECH):** This category provides recommendations for selecting, configuring, and deploying the hardware and software for emotion recognition, emphasizing safety, comfort for the child, and ease of setup. It addresses practical considerations to ensure reliable operation and minimize distractions in the intervention environment.
5. **Activities in social robot interaction (INT):** These guidelines focus on structuring and adapting robot-supported interaction activities to facilitate the natural expression of emotions and effective data acquisition. They emphasize a child-centered approach, planning flexible sessions, and varying activity types to meet therapeutic goals and sustain engagement.
6. **Symptoms data processing (PROC):** This section outlines procedures for handling and preparing collected emotional symptom data for analysis, including synchronization, managing inconsistencies, and baseline recording. It addresses the unique challenges of processing data from children with autism to ensure accurate interpretation and model development.
7. **Emotion recognition in children with autism (EMO):** These guidelines provide insights into the unique ways children with autism express emotions and how these atypical patterns impact automatic emotion recognition. They highlight the importance of understanding these differences to avoid misinterpretation and inform the development of more appropriate recognition models.
8. **Design of research studies (RES):** This category offers guidance for structuring research studies on emotion recognition in robot-supported interventions, covering aspects like group construction, data sharing, and comprehensive reporting of participant characteristics and study outcomes. It aims to enhance the validity, transparency, and reusability of research findings.
9. **Reporting studies on children with ASD (REP):** This section provides recommendations for clear, precise, and inclusive language when documenting studies, emphasizing accurate terminology for devices, channels, modalities, and emotional states. It promotes consistency and avoids ambiguity in scientific communication within the field.

During the process of drafting and evaluating the document, the following number of guidelines have been included in successive versions: 49 in version 1.0, 46 in version 1.1, and 44 in the final version 1.2. A final list of guidelines with a description of the categories is presented in Appendix A.

4 Evaluation of the Guidelines – Methodology and Criteria

A comprehensive evaluation was conducted to ensure the usability of the developed guidelines. According to the triangulation rule, we decided to apply three diverse research techniques, including:

- a questionnaire;
- focus group;
- expert evaluation using the AGREE instrument.

The first method allowed us to obtain quantitative data, the second was qualitative, and the latter provided qualitative and quantitative information.

The questionnaire survey and focus groups were carried out on version 1.0 of the guidelines and involved 48 master's and PhD students (22 females and 26 males), aged between 21 and 29 years, who participated in an international training course on the affective loop in robot–child interaction in autism therapy, organised as part of the EMBOA project. This group was selected to represent the “Researcher” and “Technology Developer” target audiences. Their evaluation focused primarily on the clarity, readability, and “perceived applicability” of the guidelines for those entering the field, rather than clinical efficacy. Additionally, the AGREE instrument evaluation of version 1.1 of the guidelines was conducted by three experts, experienced occupational therapists and researchers who were members of the EMBOA consortium.

4.1 Evaluation Criteria

There is no standard approach to evaluating technical guidelines; therefore, diverse approaches might be considered. Technical guidelines are required for the systems; thus, we might apply the quality criteria used in requirement engineering, such as the “five C” rule (clear, concise, concrete, complete, and consistent). Another approach is using the appraisal processes that are applied for the evaluation of medical guidelines. The development of medical guidelines has been studied much more than for the technical domain. For example, the AGREE guideline appraisal tool uses the following domains for evaluation: scope and purpose,

stakeholder involvement, the rigor of development, clarity of presentation, applicability, and editorial independence [19].

In our study, we decided to combine those two approaches. We have applied the AGREE II tool for the evaluation with some adaptations resulting from the non-medical domain. Additionally, we used the following criteria: understandability, adequateness of description, and applicability. Those criteria are not independent of each other. Adequate description influences understandability, and both might influence the applicability of a guideline.

As we used several research methods (questionnaire, focus groups, and expert evaluation with AGREE), we needed to further operationalise the quality criteria for each method used (see Sect. 4.5).

4.2 Questionnaire

The primary purpose of the questionnaire evaluation was to assess the understandability and completeness of the description of the developed guidelines and their applicability in practice.

Respondents were evaluating each guideline separately. The evaluation of each guideline was based on three questions, which addressed the following issues:

1. adequate description on a 5-point symmetric scale ranging from too little to too much (with a 3 being the best rating);
2. understandability of the guideline and its description – on a 5-point agree-disagree scale (with 5 being the best rating);
3. applicability of the guideline – with a 5-point agree-disagree scale (with 5 being the best grade).

A paper questionnaire, as shown in Fig. 2 was handed out to 48 participants, who were asked to read the guidelines one at a time and answer three questions about each. After the evaluation, the collected questionnaires were digitised to analyse the results.

4.3 Focus Groups

To complement the quantitative data obtained from the survey, a focus group study was conducted to gather qualitative insights and facilitate a deeper discussion on the guidelines' utility. The primary goal was to discuss the most critical recommendations and to identify specific reasons for any ambiguities found during the individual evaluation. Consequently, after completing the questionnaire, the same respondents were invited to participate in the focus groups, ensuring they were already familiar with the content. Due to

Guidelines for emotion recognition in robot-based intervention in autism

Individual Questionnaire

For each guideline: read, then evaluate according to the criteria on 5-item scale by underlining, crossing or putting in a circle.

Guideline	Criterion	Evaluation						
		Too little	1	2	3	4	5	Too much
GEN1	Guideline is described:	Too little	1	2	<u>3</u>	4	5	Too much
	Guideline is understandable:	Do not agree	1	2	3	4	<u>5</u>	Do agree
	Guideline is easy to apply:	Do not agree	1	2	3	<u>4</u>	5	Do agree
GEN2	Guideline is described:	Too little	1	2	3	4	<u>5</u>	Too much
	Guideline is understandable:	Do not agree	1	2	3	<u>4</u>	5	Do agree
	Guideline is easy to apply:	Do not agree	1	2	<u>3</u>	4	5	Do agree
GEN3	Guideline is described:	Too little	1	2	<u>3</u>	4	5	Too much
	Guideline is understandable:	Do not agree	1	2	3	4	<u>5</u>	Do agree
	Guideline is easy to apply:	Do not agree	1	2	3	4	<u>5</u>	Do agree
CH1	Guideline is described:	Too little	1	2	<u>3</u>	4	5	Too much
	Guideline is understandable:	Do not agree	1	2	3	4	<u>5</u>	Do agree
	Guideline is easy to apply:	Do not agree	1	2	3	<u>4</u>	5	Do agree

Fig. 2 Sample paper questionnaire

the large number of participants, it was decided to conduct the study in 7 parallel focus groups, with 6 to 8 participants in each group.

The focus groups were facilitated and lasted between 20 and 40 minutes. During the process, each group was asked to come to a common opinion on the following topics:

- 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?

During the discussion of individual questions, participants were asked to argue their choices, leading to a better understanding of individual guidelines and identified problems. The moderators took notes from the meeting, which, in a further step, were used to improve the guidelines.

4.4 AGREE II Instrument

The final validation stage was carried out by three experts using the AGREE II (The Appraisal of Guidelines for Research and Evaluation) instrument [19]. This step was intended to validate the development process against established standards and ensure the guidelines are robust enough for professional application.

All experts were members of the EMBOA project consortium and represented complementary expertise: (i) a Professor at a Medical Faculty specializing in the diagnosis and treatment of children with autism and serving as the National Coordinator for autism; (ii) a Professor at a Technical University with expertise in social assistive robots, human–robot interaction, and affective computing, primarily applied to education, therapy, and treatment of children with disabilities; and (iii) a Senior Researcher focusing on assistive technologies for children with autism, with a long-standing research focus on the use of social robots in autism therapy and education to foster basic communication and social interaction skills. The experts participated in an observational studies as well as brainstorming sessions.

The AGREE II is a tool for the evaluation of the process of the development of practice guidelines and the quality of

the reporting. The AGREE II consists of 23 items, divided into 6 quality domains (scope and purpose, stakeholder involvement, rigour of development, clarity of presentation, applicability, and editorial independence) and an additional 2 items of a general nature. The AGREE instrument was developed to evaluate guidelines for medical interventions but has also been used in therapeutic domains.

To adapt the tool to evaluate the presented guidelines, which are not strictly medical, the following changes were made:

- question 11 from domain 3 was excluded: The health benefits, side effects, and risks have been considered in formulating the recommendations;
- question 16 from domain 6 was excluded: The different options for management of the condition or health issue are clearly presented;
- in question 2 from domain 1 phrase “health question” was replaced with “study question”.

Experts performed evaluations in spreadsheets, which were then compiled for statistical analysis.

4.5 Operationalisation of Variables

Understandability in our study is measured in the following ways:

- in a questionnaire: a direct question on whether a reader finds a guideline understandable is asked per each of the guidelines with a 5-point symmetric agree-disagree scale (with 5 being the best grade); then, we calculate mean, standard deviation, and median score per guideline, per category of the guideline, and the whole set of guidelines;
- in focus groups: an open question is asked about the guidelines most hard to understand (qualitative information);
- in the expert evaluation: the AGREE tool has a section regarding the clarity of presentation and a detailed item on “the recommendations are specific and unambiguous”, with a 7-point Likert scale.

Adequateness of description in our study is measured in the following ways:

- in a questionnaire: a direct question on whether a reader finds a guideline described adequately or too little/too much is asked per each of the guidelines with a 5-point symmetric agree-disagree scale (with 3 being the best grade); then we calculate mean, standard deviation, and

median score per guideline, per category of the guideline, and the whole set of guidelines;

- in focus groups:
 - an open question is asked if any of the guidelines could be removed (qualitative information);
 - an open question is asked if any of the guidelines could be added (qualitative information);
- in expert evaluation: AGREE tool has several items connected to the adequateness of description, including (each with a 7-point Likert scale):
 - the overall objectives of the guideline are specifically described,
 - the questions covered by the guideline are specifically described,
 - the population (patients, public, etc.) to whom the guideline is meant to apply is specifically described,
 - the strengths and limitations of the body of evidence are clearly described,
 - the methods for formulating the recommendations are clearly described,
 - key recommendations are easily identifiable.

Applicability in our study is measured in the following ways:

- in a questionnaire: a direct question on whether a reader finds a guideline applicable is asked per each of the guidelines with a 5-point symmetric agree-disagree scale (with 3 being the best grade); then, we calculate mean, standard deviation, and median score per guideline, per category of the guideline, and the whole set of guidelines;
- in focus groups:
 - an open question is asked to list items (guidelines) that are the hardest to apply;
 - three open questions are asked to list 3 guidelines that are the most valuable for therapists, developers, and researchers;
- in the expert evaluation:
 - the AGREE tool has four items grouped under the domain of “applicability”, and we decided to use all of them (with each item evaluated on a 7-point Likert scale);
 - in the expert evaluation: the AGREE tool has a final item on “I would recommend this guideline for use” with YES - YES with modifications-NO scale

Table 3 Statistic metrics from the questionnaire for all guidelines and all metrics

Criteria	Scale	Best	Worst	Mean (SD)	Median	df	is normal	p-value	is significant
Too little- too much description	1–5	3	1, 5	3.25 (0.68)	3	2248	no	<e-15	yes
Understandability	1–5	5	1	4.57 (0.69)	5	2286	no	<e-15	yes
Applicability	1–5	5	1	3.985 (0.98)	4	2284	no	<e-15	yes

Table 4 Description adequateness metric statistics from the questionnaire for guidelines groups

Guidelines group	Mean (SD)	Median	df	is normal	$\mu < 3.5$		$\mu > 2.5$	
					p-value	is significant	p-value	is significant
GEN	3.23 (0.83)	3	140	no	4.1e-05	yes	<e-15	yes
CH	3.23 (0.76)	3	187	no	8.7e-08	yes	<e-15	yes
SYM	3.31 (0.68)	3	469	no	7.4e-14	yes	<e-15	yes
TECH	3.24 (0.72)	3	280	no	4.1e-11	yes	<e-15	yes
INT	3.28 (0.76)	3	234	no	1.7e-07	yes	<e-15	yes
PROC	3.43 (0.70)	3	328	no	0.006	yes	<e-15	yes
EMO	3.07 (0.67)	3	230	no	<e-15	yes	<e-15	yes
RES	3.13 (0.67)	3	226	no	<e-15	yes	<e-15	yes
REP	3.21 (0.63)	3	182	no	9.5e-11	yes	<e-15	yes

4.6 Data Analysis and Statistical Tests

Most of the quantitative data obtained in the study (questionnaire) are ordinal. Answers to the questions were assigned numeric values ranging from 1 to 5, with labels assigned to extremes. Some research suggests that for such variables, one can calculate mean and standard deviation when the scale is linear (same distance between the values) [54, 55]. We provided mean, standard deviation, and median values for the variables derived from the questionnaire. We calculated the Shapiro statistical test to check the normality of distributions, and (as expected) the distributions were not normal, so we decided to use a non-parametric test to check statistical significance. All of the tests were calculated using the R function.

For *Understandability* and *Applicability* variables, the scale ranged from 1 to 5 respectively, with 5 being the best and 3 being the neutral score. The test we used was a one-sample Wilcoxon Signed-Rank test, one-sided. We assumed that the value should be significantly higher than 3.5 for both variables.

For *Adequateness of description* variable, the scale ranged from 1 to 5, with 3 being the best and both 1 and 5 being the worst score. The test we used was a one-sample Wilcoxon Signed-Rank test, one-sided, but we used it twice – the first to confirm that the score was significantly lower than 3.5 and the second to verify that the score was significantly higher than 2.5.

Numerical scores for the evaluation using the AGREE II instrument were calculated independently for each domain, following the protocol outlined in the instrument's user manual [19]. Each item was rated by the three experts on a 7-point scale (ranging from 1–Strongly Disagree to 7–Strongly Agree). The *Obtained score* represents the sum

of the specific item scores within a domain across all three appraisers. The *Maximum possible score* is determined by multiplying the number of items in the domain by the number of appraisers (3) and the maximum rating (7). Similarly, the *Minimum possible score* is the product of the number of items, the number of appraisers, and the minimum rating (1). It is important to note that the number of items per domain varied due to the exclusion of non-applicable medical criteria as described in Sect. 4.4. Finally, the scaled domain score is presented as a percentage using the following formula:

$$\frac{\text{Obtained score} - \text{Minimum possible score}}{\text{Maximum possible score} - \text{Minimum possible score}} \quad (1)$$

5 Evaluation Results

The primary objective of this section is to demonstrate how the evaluation results drove the iterative refinement of the guidelines from version 1.0 to the final version 1.2. By triangulating data from questionnaires, focus groups, and expert appraisal, we identified specific areas requiring clarification and expansion. The quantitative survey data provided a baseline for the overall quality (Table 3) and helped pinpoint specific guideline categories that were rated lower in understandability or applicability (Tables 4, 5 and 6). These statistical insights were instrumental in determining which sections needed rewriting. The process culminated in a formal expert review, the results of which (Table 7) validated the methodological rigour of the final guidelines. The detailed findings are presented below, organised by the evaluation method used.

Table 5 Applicability metric statistics from the questionnaire for guidelines groups

Guidelines group	Mean (SD)	Median	df	is normal	p-value	is significant
GEN	3.95 (0.91)	4	140	no	7.8e-08	yes
CH	3.96 (0.93)	4	187	no	4.5e-10	yes
SYM	4.18 (0.92)	4	469	no	<e-15	yes
TECH	4.11 (1.02)	4	280	no	<e-15	yes
INT	4.13 (0.98)	4	234	no	<e-15	yes
PROC	3.63 (1.02)	4	328	no	0.006	yes
EMO	3.52 (1.11)	3	230	no	0.329	no
RES	4.10 (1.04)	4	226	no	4.2e-15	yes
REP	4.25 (0.97)	5	182	no	<e-15	yes

Table 6 Statistic metrics for the AGREE instrument for guidelines groups

Guidelines group	Individual experts			All experts		
	Expert 1	Expert 2	Expert 3	Overall quality	Recommendation for use	AGREE total score
	%	%	%	Max 7	Max 3	%
GEN	90.33	92.00	68.67	5.97	2.87	83.00
CH	72.50	93.00	85.75	6.20	2.90	84.25
SYM	87.00	98.70	84.20	6.45	2.96	90.10
TECH	81.50	99.67	80.00	6.13	2.82	87.83
INT	76.00	99.20	77.75	6.30	2.84	87.00
PROC	76.86	99.50	87.14	6.06	2.94	88.14
EMO	89.40	97.80	88.20	5.76	3.00	92.40
RES	82.80	99.60	91.80	5.84	2.78	93.00
REP	82.50	99.00	93.00	6.23	2.48	94.50
Mean	82.2	98.1	84.7	6.16	2.88	89.2

Table 7 Understandability metric statistics from the questionnaire for guidelines groups

Guidelines group	Mean (SD)	Median	df	is normal	p-value	is significant
GEN	4.42 (0.83)	5	140	no	<e-15	yes
CH	4.59 (0.67)	5	187	no	<e-15	yes
SYM	4.71 (0.61)	5	469	no	<e-15	yes
TECH	4.65 (0.63)	5	280	no	<e-15	yes
INT	4.54 (0.73)	5	234	no	<e-15	yes
PROC	4.42 (0.81)	5	328	no	<e-15	yes
EMO	4.48 (0.77)	5	230	no	<e-15	yes
RES	4.56 (0.72)	5	226	no	<e-15	yes
REP	4.57 (0.76)	5	182	no	<e-15	yes

5.1 Evaluation with Questionnaire

The questionnaire survey focused on collecting quantitative data to evaluate the quality of the prepared initial version of the guidelines in terms of understandability, adequateness of description, and practical applicability. It is important to note that the guideline numbering in this subsection corresponds to version 1.0 of the guidelines. Because adjustments were made to the guidelines that received insufficient scores, based on the analysis of the data collected, the numbering does not necessarily correspond to the numbering of the guidelines in the Appendix A.

To identify guidelines with insufficient adequateness of description, a threshold of a rating of 1 or 2 by at least 10

respondents was employed as a criterion. This led to the identification of five guidelines (GEN3, CH1, INT2, EMO4, and RES3) falling into this category. On the other hand, 8 guidelines (GEN2, CH3, SYM1, SYM10, TECH5, INT1, PROC2, and PROC6) received an average score above 3.5, indicating that they may have been too detailed.

In terms of understandability rating, only one guideline scored lower than 4, and 14 scored below 4.5. Finally, the lowest applicability scores were obtained by 3 guidelines, all of which also had relatively low understandability.

These identified guidelines have been thoroughly reviewed, and their description has been revised or merged with others.

5.2 Evaluation with Focus Groups

The focus groups proved to be a valuable source of qualitative information beyond simply identifying which guidelines needed to be changed. They provided deep insights into how the description could be improved.

During the discussion, a number of guidelines were identified that were repetitive in scope or that were in contradiction to one another. As a result, focus group participants suggested merging some of them.

The least understood guidelines were also identified, with PROC2 being mentioned by 4 groups, PROC5 and GEN3 by 3 groups, and RES3, SYM10, TECH3, PROC7, SYM1, INT1, REP2, and INT5 by one group.

Similarly, in terms of potential applicability issues, all EMO guidelines were indicated by 3 focus groups, PROC2, PROC5, RES2, and REP4 by 2, and CH4, PROC3, PROC6, PROC7, TECH4, SYM3, SYM6, SYM9, REP3, and RES3 by one.

Finally, participants suggested adding page numbering to the guidelines.

Based on the results of the focus groups and questionnaires, the guidelines have been refined to version 1.1.

5.3 Evaluation with the AGREE Instrument

The results of the evaluation of version 1.1 of the guidelines using the AGREE II tool are presented in Tables 6 and 7, which shows both the individual expert scores and the aggregated score.

A procedure for updating the guidelines received the lowest score, with 4.95 out of a possible 7 points. The description of evaluation criteria (5.02) and the description of applicability (5.73) also scored low.

Among the qualitative remarks, the most important were:

- missing a statement that the funding body did not influence the content of the guidelines;
- adding some references to improve the body of knowledge visibility in guidelines;
- the 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 on what to do.

The feedback obtained through the AGREE II instrument led to the final refinement of the document into version 1.2, which is provided in Appendix A. It should be noted that while version 1.2 was not subjected to a new round of formal evaluation, the modifications made were exclusively editorial and technical in nature, responding directly

to expert requests (e.g., adding citations and transparency statements). As the core substance and specific recommendations remained unchanged from version 1.1, the high scores for understandability, adequateness, and applicability reported in the preceding sections are considered valid for the final version.

6 Discussion

6.1 Summary of Quantitative Results

The objective of the paper was formulated as follows: **the proposed guidelines are understandable, adequately described, and applicable**, and for each of the criteria a set of measures was proposed (see Sect. 4.5 on variables operationalisation).

6.1.1 Understandability

Concerning the understandability measure in the questionnaire, the overall mean score for the guidelines set was 4.57 (on a scale from 1 to 5, with 5 being the best score), and the result was found statistically significant. We also measured the mean score per category (see Tables 7 and 5), and the mean score differed from 4.42 for the GEN and PROC categories to 4.71 per SYM category of the guidelines, with all results for categories being statistically significant.

In the focus groups, only three out of over forty guidelines were mentioned more than once as being the least understandable. Moreover, eight guidelines were mentioned by only one of the seven focus groups in this context. All of them were revised after obtaining this score.

In the evaluation with the AGREE instrument, the item on “the recommendations being specific and unambiguous” received a mean of 6.56 score (on a scale from 1 to 7, with 7 being the best score).

Considering the above results from the three evaluation methods, we conclude that the guidelines were understandable.

6.1.2 Adequateness of Description

Concerning the adequateness of the description measure in the questionnaire, the overall mean score for the guidelines set was 3.25 (on a scale from 1 to 5, with 3 being the best score), and the result was found statistically significant. We also measured the mean score per category (see Table 4), and the mean score differed from 3.07 for EMO category to 3.43 for PROC category of guidelines, with all category results being statistically significant.

In the focus groups, none of the groups suggested the removal of any guideline – they instead suggested merging them, which was performed according to the suggestions.

In the evaluation with the AGREE instrument, there were six items related to the adequateness of description criteria, and the scores were as follows (on a scale from 1 to 7, with 7 being the best score):

- the item *the overall objectives of the guideline are specifically described* was scored 6.51,
- the item *the questions covered by the guideline are specifically described* was scored 6.64,
- the item *the population to whom the guideline is meant to apply is specifically described* was scored 6.16,
- the item *the strengths and limitations of the body of evidence are clearly described* was scored 6.07,
- the item *the methods for formulating the recommendations are clearly described* was scored 6.05,
- the item *key recommendations are easily identifiable* was scored 6.59.

Taking into account the above results from the three evaluation methods, we conclude that the guidelines were found adequately described.

6.1.3 Applicability

When considering the applicability of the guidelines, it is important to be aware that certain challenges may result in a less straightforward application of the guidelines. Examining the questionnaire scores, it is clear that guidelines within the GEN, PROC, EMO, and RES sections, as well as specific guidelines within the TECH and CH sections, received lower scores for their applicability criteria.

The GEN section consists of general recommendations intended as preliminary guidance, which makes it difficult to directly apply them due to the following challenges:

- the complexity of describing and detecting emotional phenomena, particularly when using automatic recognition technologies to classify affect;
- working with children on the autism spectrum with their deficits, distressing behaviours;
- different emotional expressions compared to typically developing peers;
- both therapeutic interventions and research studies involving children with autism require a child-centred approach, which is critical for the well-being of the child, but can present limitations in terms of research due to potential data gaps, lower quality data, and limited reproducibility.

Despite our concerted efforts to enhance the applicability and understandability of the guidelines, certain challenges remain, such as obtaining accurate ground truth for a child's emotional state, ensuring complete silence in a research environment, and establishing publicly accessible datasets.

It is important to distinguish between the types of applicability assessed by the different evaluator groups. The quantitative scores from the student cohort (questionnaire) primarily reflect “perceived applicability” – how easily a novice researcher believes they could implement the guideline based on the text. This validates the guidelines as an accessible training tool. Conversely, the “clinical applicability” and validity of the recommendations in a therapeutic context rely on the high scores obtained in the AGREE II evaluation, which was conducted exclusively by domain expert.

6.2 Qualitative Results – Guidance for Target Groups

The guidelines have been developed for three target groups: autism therapists, technology developers, and researchers. We are aware that addressing such diverse audiences may present challenges to the guidelines' effective application, as what is valuable to one group may not be as relevant to the others. Therefore, during the focus group evaluation, participants were asked to identify the guidelines that would be most valuable to each specific target group. As a result of analysing the discussion and responses, the following audience-specific guidelines emerged.

For therapists, all of the guidelines in the SYM section were recommended to familiarise them with how emotion symptoms are captured. Similarly, the entire INT section was valuable in guiding planning and conducting interactions. Furthermore, in cases where technical support is lacking, the guidelines in the TECH section serve as valuable resources for addressing technical aspects related to observational studies. Finally, General Guideline GEN2 was considered useful for therapists to emphasise the importance of defining the study's goal.

The most relevant guidelines for technology developers are those in the PROC and EMO sections. The former provides information on processing the data, and the latter describes the specificity of the emotional symptoms expressed by children with autism. In addition, GEN1 was selected to raise awareness of the importance of the child's well-being and therapeutic purpose, and TECH2–TECH5 to provide important technical requirements for the technologies being developed. Finally, guidelines on the data obtained from the channels, i. e., CH2, CH3, were also considered useful for technology developers.

For the third target group, researchers, all the guidelines in the RES section were considered valuable. They guide how to handle studies on robot-child with autism interaction. Also, all guidelines in the REP section that describe how to report the studies are important for this target group. In addition, selected guidelines in the CH and EMO sections should be studied to understand the limitations of working with children with autism and to be able to evaluate data quality. Finally, the participants pointed out the guidelines for the baseline recording (PROC5) and the use of general-purpose models for children with autism (PROC7).

6.3 Revisions of Guideline

Having evaluated the guidelines with the three methods explained above, we have made revisions according to the results. The first version (1.0) was considered with both questionnaires and focus groups and then, the improved version (1.1) went under expert evaluation with the AGREE instrument, forming the final version (1.2), which is provided as an attachment to this paper. Major changes to the final guidelines document included merging the number of guidelines, removing repetitions and contradictions, and improving descriptions.

6.4 Validity Threats

Two types of validity threats are related to the contribution of this study – the guidelines for emotion recognition in robot-supported interventions in autism. The first group of validity threats is related to the methods used to formulate the guidelines. We performed two rounds of systematic literature reviews and two rounds of observational studies conducted on an international group of participants (children with autism interacting with a robot). The literature studies allowed us to elicit the practices used in such studies, with some remarks on their usability in this particular context (robot-based therapy). Then, the best-identified practices were used in the two observational studies. We tried to keep the observational studies replicable and follow the same process in all countries. This results in higher internal validity of the study (we can compare and analyse differences between the countries) but lower generalisability (external validity) – only selected practices were covered (the most relevant according to the literature review), and only one robot was used (Kaspar). There is always a trade-off between internal and external validity, and in this study, we decided to choose the internal one as a priority. We added a generalisability statement for each guideline to demonstrate which guidelines are considered robot-agnostic and which are specific for Kaspar or similar robots.

Another group of validity threats is related to the evaluation study. Evaluation of guidelines is a complex task. First, guidelines are not procedures or processes unambiguously defined or divided into precise steps. As guidelines are more general, it is more difficult to define quality attributes to evaluate them. Out of multiple dimensions of guidelines evaluation, we have selected certain attributes to be the most important regarding the usability of the proposed guidelines. In the questionnaire, we focused on adequateness of description, understandability, and applicability as they seemed to be the most important ones, due to the following reasons: Guidelines are only as good as they are adequately described and can be understood by target users, while practical use of a guideline depends on applicability. To mitigate this validity threat, we also applied the AGREE II instrument as specified above (we excluded only two items referring to medical conditions), and the instrument covers more evaluation dimensions, which was also reported. Moreover, in the focus group study, the participants were also asked open questions, which allowed us to gain qualitative feedback on issues not covered by the questionnaire.

Another validity threat is related to the choice of participants in the evaluation process. The group invited to participate in the survey and focus group study had a mixed background – researchers, occupational therapists, and participants from the technical domain – but was not balanced, as those from the technical domain were the most numerous. All of these groups are among the target audiences of the guidelines; however, the study would benefit from a more balanced composition. Moreover, we did not ask participants to report their background along with the questionnaire, which may be considered a limitation of the study. In addition, a substantial proportion of the survey and focus group participants were newly trained and lacked meaningful experience in the therapy of children with ASD. We aimed to keep the group relatively large to reduce the influence of individual differences or temporary factors (e.g., disposition on the day of participation) on the evaluation scores; however, the confounding effect of participant background was not controlled for in this study.

Furthermore, the evaluation using the AGREE II instrument involved only three experts, all of whom had previously participated in the observational studies and brainstorming sessions. This limited number of evaluators and their prior involvement may have introduced an additional source of bias.

Finally, a potential threat to external validity may arise from the reliance on a single robot platform (Kaspar) during the observational studies. Although the observational studies relied on the humanoid robot Kaspar, the proposed guidelines were derived from broad literature reviews and are intended to be adaptable to various robot morphologies,

as categorised by Scassellati et al. [56] and Diehl et al. [57]. For humanoid robots, guidelines regarding familiarisation (INT2) are critical to manage potential over-stimulation or the “uncanny valley” effect. Conversely, for non-humanoid or animal-like robots (e.g., Paro), where interaction is often tactile rather than face-to-face, the focus shifts from robot-embedded sensors to external environmental monitoring (SYM2) to avoid data loss during hugging or handling. For mobile or ball-shaped robots (e.g., Leka), the primary challenge lies in maintaining consistent observation channels (CH1, CH2) while the device and child are in motion. In these scenarios, robot-mounted sensors are often unreliable for continuous face tracking. Therefore, the guidelines advise adapting the setup to prioritize wide-angle external cameras or wearable biosensors (TECH3) over robot-centric data capture. While the specific technical implementation must adapt to the robot’s form factor, the core methodological guidelines regarding study design (RES) and reporting (REP) remain universally applicable across platforms.

To the authors’ knowledge, no such set of guidelines has been proposed for the combination of social robots and emotion recognition technologies. The objective was to establish a bridge from the technological capabilities of robots and emotion recognition to their practical applications and studies. There are specific decision points and trade-offs that only emerge when emotion recognition and robotics are combined, and guidelines on observational channels (CH) and interaction (INT) specifically address those issues.

7 Conclusion and Future Works

The article has presented the development and subsequent evaluation of comprehensive guidelines for automatic emotion recognition in robot-assisted interventions for children with autism. A thorough analysis of the evaluation results has supported the initial hypothesis, confirming that the guidelines developed are clear and comprehensible, adequately described, and crucially applicable in practice. It should be noted that while the final evaluation was conducted on version 1.1, the results were used to finalize version 1.2 by incorporating minor editorial refinements; as the core substance remained unchanged, the evaluation findings are representative of the final set of recommendations.

The developed guidelines show promise as a foundation for subsequent research efforts. They provide a robust platform for further exploration of extending human-robot interaction in therapeutic contexts for children with autism, particularly by incorporating affective features. This potential extension of interventions could mark a significant advance in the field of social skills therapies, offering new ways to improve the effectiveness of therapeutic approaches

for children with autism. In addition to increased effectiveness of interventions, the application of the guidelines will likely increase the children’s enjoyment of their therapeutic sessions. A robot equipped with a robust emotion recognition technology will manage to adapt in real time to the current needs and preferences of the children, e.g., by automatically interspersing motivating or playful scenarios whenever a child gets bored or frustrated.

While the guidelines were developed specifically for autism interventions, many recommendations have significant transferability to related domains. The technical guidelines regarding observational channels (CH), symptom acquisition (SYM), and data processing (PROC) are directly applicable to the broader field of educational robotics and Socially Assistive Robotics (SAR) for other neurodivergent populations or elderly care. In any context where a social robot is used to engage vulnerable users, the principles of non-intrusive sensing, emotional safety, and child-centered interaction design presented here provide a robust framework for establishing effective affective loops.

Future research will focus on developing the affective loop approach to social robot therapy for children with autism, which allows for the seamless integration of emotion recognition and elicitation. Through this research, we expect to support researchers working in this field, and to propose innovative ways to advance therapeutic techniques and empower children with autism to navigate their social world with greater confidence and skill.

A Guidelines for Emotion Recognition in Robot-Supported Interventions in Autism

A.1 General Guidelines (GEN)

A.1.1 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 focus on a child and the needs of the therapy. The child and therapy should not 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 [12]:

- For what purpose are emotions recognised – 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?

Generalisability statement: This guideline is considered robot-agnostic.

A.1.2 Guideline GEN2: Start with What You Want to Know About the 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.

Psychological research has no unambiguous definition of human emotion [58]. However, the concept of discrete emotions is widely accepted. There are multiple models to represent emotions [59]. Analysing papers dealing with emotion recognition, we observed the issue of distinguishing and analysing emotions. Most of the papers use two emotion models: Ekman's basic emotions (joy, anger, fear, disgust, sadness, surprise) [60] 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. Apart from emotions, some studies 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 [31].

Sometimes, the six basic emotions (happy, scared, angry, surprised, disgusted, sad) are good 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 [61, 62].

Generalisability statement: This guideline is considered robot-agnostic.

A.1.3 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 GDPR (<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. In addition, other constraints arising from legal aspects, such as those arising from policies being developed to address the use of AI techniques, such as the European Union's "The Artificial Intelligence Act", need to be taken into account.

Generalisability statement: This guideline is considered robot-agnostic.

A.2 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 into emotion symptoms, can be obtained from multiple channels. A channel is a medium for recording a signal, e.g., 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 [31]:

- 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.

A.2.1 Guideline CH1: While Choosing an Observational Channel Consider the 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 also consider the context (including place, timing) of the study being conducted.

Generalisability statement: This guideline is considered robot-agnostic.

A.2.2 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 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 partially cover the face. Rapid hand movements or fidgeting may cause erroneous readings when collecting physiological data with a wristband. These are situations that cannot be avoided and are inherent when recording children.

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

Generalisability statement: This guideline is considered robot-agnostic.

A.2.3 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 between multi-channel and multimodal observation. The first one refers to the number of devices used for capturing observable symptoms, e. g., using a combination of a camera and a microphone and multiple cameras. The second one refers to analysing diverse symptoms, e. g., facial expressions and speech prosody. 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 that at least one will provide valuable information on emotional symptoms. Moreover, multimodal emotion recognition systems offer higher classification accuracy than single-modality-based solutions [63]. Therefore, it is advisable to use multiple channels during observations to collect different modalities if feasible.

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

Generalisability statement: This guideline is considered robot-agnostic.

A.2.4 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 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 [64], meaning 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 know 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

[31]. 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 distracting microphones on the table in front of the child or the child's clothes.

An example of a minimal set of devices providing 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 [31].

Generalisability statement: This guideline is considered robot-agnostic.

A.3 Reliable Acquisition of Emotional Symptoms (SYM)

The second challenge we have identified is that even though one will record the entire session using devices, the data will not contain 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, which apply to specific types of modalities obtained.

A.3.1 Guideline SYM1: Adjust the Distance Between a Child's Face and a Camera to the Resolution of the Camera

To recognise emotions from facial expressions, the whole child's face must be visible as a fraction of the obtained video image. 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 1–1.5 m from the camera's lens. The distance might be adjusted depending on camera resolution and angle. Telephoto 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. 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 for 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 (impossible to eliminate) might record only a fraction of a face, hindering 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 the size of the child's face concerning the total image before recording to get better recognition results.

Generalisability statement: This guideline is considered robot-agnostic.

A.3.2 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 is also prone to some risks. For example, when a child moves around the room or turns sideways while 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 insufficient face image quality. Angular or tilted views might make facial expressions hard to analyse as well. When using multiple cameras, pay attention to synchronisation challenges (see PROC1 and PROC2).

The effectiveness of emotion recognition based on facial expression analysis is significantly affected by the camera's position relative to the face [65]. 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.

Additional side cameras can capture behavioural patterns such as posture and gestures if multiple cameras are used.

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

Generalisability statement: This guideline is considered partially robot-agnostic. While the importance of camera placement applies across different systems, mobile robots may require additional cameras compared to more static robots to reliably keep the child within the camera's field of view.

A.3.3 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.

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 consider the likely low accuracy of emotion recognition in this modality when processing the collected data.

Generalisability statement: This guideline is considered robot-agnostic.

A.3.4 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 recognising emotions based on an expression. On the other hand, poor lighting will cause underexposure, which involves the loss of shadows that also allow the 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).

Generalisability statement: This guideline is considered robot-agnostic.

A.3.5 Guideline SYM5: Choose a Silent Room – Control the Level of the Outside Noise and Limit the Noise Generated by the Room Interior

To obtain high-quality voice data from 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 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.

Avoid as much sound as possible from chairs, tables, doors, furniture, and windows. 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 everyone 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.

Generalisability statement: This guideline is considered robot-agnostic.

A.3.6 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 [66]. However, the child's voice detection is hampered if the child's voice co-occurs with other people's voices. Therefore, we recommend reducing the number of people 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.

Generalisability statement: This guideline is considered robot-agnostic.

A.3.7 Guideline SYM7: Adjust the 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 compared 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 the robot can process the child's voice.

Generalisability statement: This guideline is considered robot-agnostic.

A.4 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.

A.4.1 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. Do not place them next to children to avoid tripping and destroying the measurement environment by pulling the cable if possible. 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 cardboard box, and cables can be duct-taped to the floor. Carpets might also help to hide cables on the floor.

Generalisability statement: This guideline is not entirely robot-agnostic because it addresses the physical safety of the robot and other study equipment. Although the general idea of providing a safe and minimally distracting environment applies to all types of robots, the specific recommendations regarding robot stability and cable management are strongly dependent on the type of robot and recording

devices that are used. These factors are for example less critical for smaller or simpler robotic devices.

A.4.2 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, in round 1, the lapel microphone did not add more information on vocalisations than the internal microphone, 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) 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 and test the device in a practical setting before starting a session with a child. It is hard to calibrate the device, and for most 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 [67].

Generalisability statement: This guideline is considered robot-agnostic.

A.4.3 Guideline TECH3: For Devices That Must Be Put on a Child (On Hands, Attached to Clothing), Consider a Familiarisation Stage

Consider that some children may feel uncomfortable with some 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 we used in our studies), choose something cable-free. Wristbands are a good option as they seem like a watch and might be easily accepted by the child. 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 assembling similar wristbands (or dummies) for the robot (e. g., Kaspar), the therapist, and the caregiver, so wristbands become something special to that specific context.

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

Generalisability statement: This guideline is considered robot-agnostic. The recommendations are broadly applicable to settings that involve children wearing recording devices. The robot Kaspar is mentioned only as an example.

A.4.4 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. Performing a test session before the observation sessions (without a child) is advisable.

Generalisability statement: This guideline is considered robot-agnostic.

A.5 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 allow 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. [12]

A.5.1 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 [12]:

- 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 learnt 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 consider 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. Accurate emotion recognition is not always needed.

Automatic emotion recognition technologies are especially important when emotional skills are trained. The emotions recognised might be adjusted to the training purpose

and the child's ability. For example, only "happy" expression might be tracked during the smile training.

On the other hand, it might be interesting to capture a child's general mood or attitude 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 recognise if a child is crying or not or just distinguish between positive and negative reactions.

Generalisability statement: This guideline is considered robot-agnostic.

A.5.2 Guideline INT2: Plan a 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 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 avoid interaction at any point (see GEN3). The therapist/researcher should encourage and motivate the child to interact with the robot, not press the child.

Generalisability statement: This guideline is considered robot-agnostic.

A.5.3 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., having a vague plan in advance is good, 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 take into account that ASD is a very heterogeneous disorder. Therefore, the intervention should be adjusted to the child's attention span and 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.

Generalisability statement: This guideline is considered robot-agnostic.

A.5.4 Guideline INT4: Consider Diverse Activity Types and Diverse Difficulty Levels

Preparing different activities for specific ages, verbal and cognitive capacities, attention span, individual interests, and therapeutic goals is recommended. Moreover, it is useful to prepare activities for diverse difficulty levels to provide the best suitable intervention for each child with ASD. 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 children's interests should be considered. Many children, for example, seem to enjoy singing the most. Therefore, singing a song could be chosen as an icebreaker at the beginning of the intervention session and/or in between to relax the child whenever he or she seems stressed. On the other hand, it might be the case that singing a song is not helpful for children of all ages. Therefore, it would be helpful to find out the interests of a specific child before 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.

Generalisability statement: This guideline is partially robot-agnostic. While the principle of adapting activity types and difficulty levels to individual child characteristics is broadly applicable, the concrete activities that can be offered depend on the capabilities of the robot (e.g., singing a song). The guideline is therefore not specific to Kaspar, but to robots with comparable capabilities.

A.5.5 Guideline INT5: Perform Multiple Sessions with a Child

To evaluate the success of the robot-supported intervention compared to standard therapeutic approaches, it is highly recommended to follow the child over longer periods, i.e., to perform multiple robot interaction sessions with a child. Recruiting children performing in longitudinal studies may be challenging 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.

Generalisability statement: This guideline is considered robot-agnostic.

A.6 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.

A.6.1 Guideline PROC1: If Recording Multiple Modalities, Pay Attention to Synchronisation

Multimodal emotion recognition systems offer higher classification accuracy than single modality-based solutions [63]. 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 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 must enable channel synchronisation when pre-processing the collected data.

Generalisability statement: This guideline is considered robot-agnostic.

A.6.2 Guideline PROC2: If Recording Multiple Modalities, Pay Attention to Inconsistencies

When using multimodal 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, integrating 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 based on diverse and the same input channel analysis [68]. For example, in an experiment on an educational management game, a significant discrepancy between the self-evaluated and detected emotional states was reported [69]. In another study, we observed differences in the recognised emotional state based on facial expressions

recorded with two cameras: placed below and above the monitor [65].

Analysing the same moment of observation with the same person and context, diverse algorithms might provide different emotion recognition results. The recommended approach is evaluating channels' temporal unavailability and 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.

Generalisability statement: This guideline is considered robot-agnostic.

A.6.3 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 result in misinterpretation. Therefore, it is crucial to be quite scrupulous in noting those events. 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., something has fallen). Note down atypical behaviours of a child and events related to recording devices, such as a dead battery, a 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 interpreting the results and might be valuable from the perspective of publishing a dataset for future research (see RES2). To properly annotate, consider precise definitions of events, states and tasks to annotate, annotation by multiple people (including consistency check), and repeatable methods (tools) for the process.

Generalisability statement: This guideline is considered robot-agnostic.

A.6.4 Guideline PROC4: For the Voice Channel, the Child's Speech Must Be Distinguished from Other Sounds First

To recognise the child's emotional state from his or her voice, we first need to determine the segments with the child's vocalisations. 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 articulated child's speech with no or little background noise. 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, providing a wide range of training material for the voice activity detector is useful. 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 [70]. It should also be considered that children of different ages have different voice characteristics.

When planning a study, consider 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 the 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).

Generalisability statement: This guideline is considered robot-agnostic.

A.6.5 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 used for comparison purposes. For example, a sample of an undisturbed child's voice might be recorded for a voice channel. 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 [71] 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 [72], 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 and other physiological signals.

Similarly, for a voice channel, you may consider collecting a sample from the child before 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.

Generalisability statement: This guideline is partially robot-agnostic. While the general principle of recording children's baselines is broadly applicable, some of the suggested elicitation strategies (e. g., imitation-based prompts) depend on the capabilities of the robot. The guideline is therefore not specific to Kaspar, but to robots with comparable capabilities.

A.6.6 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, diverse methods are 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 general emotion recognition, not only autism. An individual's internal emotional state (ground truth) is hard to determine continuously, even by himself/herself. Several strategies might be used for labelling the data. In the study [73], 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 [74] and compared to therapists' reports with a consistency of approximately 83%. In a single study of human-robot interaction, loop [33] self-report was used; however, it was 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 to enhance the reliability of tagging, a clinical observer and a caregiver who knows the participant shall be included in the study [33].

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.

Generalisability statement: This guideline is considered robot-agnostic.

A.6.7 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., helps optimise existing emotion recognition models.

Generalisability statement: This guideline is considered robot-agnostic.

A.7 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.

A.7.1 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 [75]. 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.

A.7.2 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 compared to neurotypical children [76]. Children with ASD produce the same emotional facial expressions and speech at the same intensity levels as typically developing children, but facial and vocal expressions are less coordinated. The literature review summarised in the same paper shows 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 [77], 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).

Generalisability statement: This guideline is considered robot-agnostic.

A.7.3 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 [78, 79]. 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. [66], 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.

Generalisability statement: This guideline is partially robot-agnostic. Although the referenced study [66] involved interactions with the social robot Kaspar, similar results are expected with other social robots.

A.7.4 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 conversation. The action refers not only to an eye gaze but also to head-turning. This might influence capturing facial expressions and areas of interest from the eye gaze.

As the child's condition is given, and we cannot influence it, the recommendation is just to report such behaviours.

Generalisability statement: This guideline is considered robot-agnostic.

A.7.5 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, all of the life activities and modalities used as proxy symptoms for emotion recognition 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 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 [68]. Some of the emotional

expressions might be changed purposefully by a person. For example, children with autism often smile when left alone, not because they feel joy.

Therefore, the 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 a person's emotional state, considering that we could observe and measure symptoms only. Automated solutions should keep the human in the loop.

Generalisability statement: This guideline is considered robot-agnostic.

A.8 Design of Research Studies (RES)

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

A.8.1 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 since 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 [31].

Generalisability statement: This guideline is considered robot-agnostic.

A.8.2 Guideline RES2: Create an Openly Available and Well Described Dataset for Future Studies

The affective computing discipline benefits from many recorded and openly shared datasets among researchers. The sets contain anonymised feature sets or raw data, followed by emotion-related labels or values to classify the emotions. However, only a single dataset is openly available for emotion recognition applied to children with autism. The deENIGMA project created it and it might be used in future research [61]. It contains recordings from 35 children with autism.

While preparing and publishing a dataset, following the FAIR rules is advisable for open science [80]. FAIR is

an acronym for the following dataset features: Findable – Accessible – Interoperable – Reusable. 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 a globally unique and persistent identifier and 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).

Generalisability statement: This guideline is considered robot-agnostic.

A.8.3 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 generalisability 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.

Generalisability statement: This guideline is considered robot-agnostic.

A.8.4 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 – from technical problems (such as disconnection of the sensor, device failure or low battery), 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 another 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.

Generalisability statement: This guideline is considered robot-agnostic.

A.8.5 Guideline RES5: If Evoking Emotions in Your Study, Choose Appropriate Stimuli

Some research studies might require evoking emotional states rather than observing natural ones. The studies used multiple strategies 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 [81]. The most common stimuli used in the studies included pictures and video resources. However, a study [72] 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 from videos containing human faces. Although video stimulus was the most frequently used one, other studies tried: a serious game approach [82], computer-based intervention tools [83], or observation of natural human-robot interaction [33, 84].

Generalisability statement: This guideline is considered largely robot-agnostic. The suitability of emotional stimuli for children with autism primarily depends on participant characteristics and stimulus modality rather than on the specific social robot or platform used.

A.9 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.

A.9.1 Guideline REP1: Distinguish the Two Meanings of the “Emotion Recognition” Phrase

In the context of emotion recognition technologies supporting therapy of children with autism, the phrase “emotion recognition” has two meanings: emotion recognition by children with autism and recognising emotions of children with autism. The first one refers to the ability of children to recognise emotion in others. The latter concerns automatic

emotion recognition technologies applied to recognise and analyse the emotional states of a child.

Generalisability statement: This guideline is considered robot-agnostic.

A.9.2 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 emotion recognition process is analysed concerning 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, and physiological activities: heart activity, unconscious muscle activity, respiration, and thermal regulation. In addition, activating 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 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.

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

Generalisability statement: This guideline is considered robot-agnostic.

A.9.3 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 to not imply that one individual is superior to another in 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 [12, 31]. We suggest using more appropriate alternatives such as “typically developing children” or “neurotypical”.

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”.

Generalisability statement: This guideline is considered robot-agnostic.

A.9.4 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 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 recognised state and defining it properly. For example, if you recognise 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 [83, 85, 86] 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 a child's meltdown crisis. Appendix A as psychologists define particular discrete emotions differently, it is recommended to choose one of those definitions for your study and report it.

Generalisability statement: This guideline is considered robot-agnostic.

Acknowledgements We want to thank all families, therapists, and researchers who participated in our study.

Authors Contribution All authors contributed to the conception of the study and the design of the observational studies, as well as the preparation of the guidelines. Ben Robins, Tatjana Zorcec, Hatice Kose, Duygun Erol Barkana, Agnieszka Landowska, and Michal R. Wrobel conducted the field studies. Michal R. Wrobel and Agnieszka Landowska designed the evaluation study, facilitated the survey and focus groups, and performed the statistical analysis. Tatjana Zorcec, Hatice Kose, and Ben Robins evaluated the guidelines using the AGREE II instrument. Michal R. Wrobel, Agnieszka Landowska, Katrin D. Bartl-Pokorny, Tatjana Zorec, Duygun Erol Barkana, and Hatice Kose wrote the first draft of the manuscript. All authors commented on previous versions and contributed to the final version. All authors read and approved the final manuscript. Michal R. Wrobel coordinated the work.

Funding This work was supported by the European Commission's Erasmus+ Project (EMBOA, Affective loop in socially assistive robotics as an intervention tool for children with autism) under Contract 2019-1-PL01-KA203-065096. The European Commission's support

for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use that may be made of the information contained therein.

Data Availability Data collected during the observation sessions is available upon request.

Declarations

Ethical Approval The observational studies were conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethical Committee of Gdansk University of Technology.

Competing Interests The authors declare no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Maenner MJ, Warren Z, Williams AR, Amoakohene E, Bakian AV, Bilder DA, Durkin MS, Fitzgerald RT, Furnier SM, Hughes MM et al (2023) Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2020. *MMWR Surveill Summ* 72(2):1. <https://doi.org/10.15585/mmwr.ss7202a1>
- Zeidan J, Fombonne E, Scorah J, Ibrahim A, Durkin MS, Saxena S, Yusuf A, Shih A, Elsabbagh M (2022) Global prevalence of autism: a systematic review update. *Autism Res* 15(5):778–790. <https://doi.org/10.1002/aur.2696>
- Norrelgen F, Fernell E, Eriksson M, Hedvall Å, Persson C, Sjölin M, Gillberg C, Kjellmer L (2015) Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism* 19(8):934–943. <https://doi.org/10.1177/1362361314556782>
- Tager-Flusberg H, Kasari C (2013) Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. *Autism Res* 6(6(6)):468–478. <https://doi.org/10.1002/aur.1329>
- Sandbank M, Bottema-Beutel K, LaPoint SC, Feldman JI, Barrett DJ, Caldwell N, Dunham K, Crank J, Albarran S, Woynaroski T (2023) Autism intervention meta-analysis of early childhood studies (project aim): updated systematic review and secondary analysis. *BMJ* 383
- Clarke EB, Lord C (2023) Social competence as a predictor of adult outcomes in autism spectrum disorder. *Dev Psychopathol* 36(3):1442–1457. <https://doi.org/10.1017/S0954579423000664>
- Hodges H, Fealko C, Soares N (2020) Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Transl Pediatr* 9(Suppl S1):S55–S65. <https://doi.org/10.21037/tp.2019.09.09>
- Banire B, Al Thani D, Qaraq M (2023) One size does not fit all: detecting attention in children with autism using machine learning. *User Modeling User-Adapted Interact* 1–33. <https://doi.org/10.1007/s11257-023-09371-0>
- Wang LA, Petrulla V, Zampella CJ, Waller R, Schultz RT (2022) Gross motor impairment and its relation to social skills in autism spectrum disorder: a systematic review and two meta-analyses. *Psychological Bull* 148(3–4):273. <https://doi.org/10.1037/bul0000358>
- Akin-Bulbul I, Ozdemir S (2023) Imitation performance in children with autism and the role of visual attention in imitation. *J Autism Dev Disord* 53(12):4604–4617. <https://doi.org/10.1007/s10803-022-05726-5>
- Schadenberg BR, Reidsma D, Evers V, Davison DP, Li JJ, Heylen DK, Neves C, Alvito P, Shen J, Pantić M, Schuller BW, Cummins N, Olaru V, Sminchisescu C, Dimitrijević SB, Petrović S, Baranger A, Williams A, Alcorn AM, Pellicano E (2021) Predictable robots for autistic children—variance in robot behaviour, idiosyncrasies in autistic children's characteristics, and child-robot engagement. *ACM Trans Comput-Hum Interact* 28(5):1–42. <https://doi.org/10.1145/3468849>
- Bartl Pokorny KD, Pykala M, Uluer P, Barkana DE, Baird A, Kose H, Zorcec T, Robins B, Schuller BW, Landowska A (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>
- Salimi Z, Jenabi E, Bashirian S (2021) Are social robots ready yet to be used in care and therapy of autism spectrum disorder: a systematic review of randomized controlled trials. *Neurosci. Biobehav. Rev* 129:1–16. <https://doi.org/10.1016/j.neubiorev.2021.04.009>
- Dubois-Sage M, Jacquet B, Jamet F, Baratgin J (2024) People with autism spectrum disorder could interact more easily with a robot than with a human: reasons and limits. *Behavioral Sci* 14(2):131. <https://doi.org/10.3390/bs14020131>
- Herrero-Martín J, Fonseca D, Caro-Via S, Canaleta X (2024) Development of personalized profiles of students with autism spectrum disorder for interactive interventions with robots to enhance language and social skills. *Front. Psychiatry* 15, 1455627 (<https://doi.org/10.3389/fpsy.2024.1455627>)
- Kouroupa A, Laws KR, Irvine K, Mengoni SE, Baird A, Sharma S (2022) The use of social robots with children and young people on the autism spectrum: a systematic review and meta-analysis. *PLoS One* 17(6), 0269800 (<https://doi.org/10.1371/journal.pone.0269800>)
- Chung EY-H, Kuen-Fung Sin K, Chow, D.H.-K. (2025) Effectiveness of robotic intervention on improving social development and participation of children with autism spectrum disorder—a randomised controlled trial. *J Autism Dev Disord* 55(2):449–456. <https://doi.org/10.1007/s10803-024-06236-2>
- Diehl JJ, Schmitt LM, Villano M, Crowell CR (2012) The clinical use of robots for individuals with autism spectrum disorders: a critical review. *Res autism spectr Disord*. Jan 6(1). <https://doi.org/10.1016/j.rasd.2011.05.006>
- Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE et al (2010) Agree II: advancing guideline development, reporting and evaluation in health care. *Can Med Assoc J* 182(18):E839–E842. <https://doi.org/10.1503/cmaj.090449>
- Alabdulkareem A, Alhakhani N, A-N (2022) A.: a systematic review of research on robot-assisted therapy for children with autism. *Sensors* 22(3):944. <https://doi.org/10.3390/s22030944>

21. Santos L, Annunziata S, Geminiani A, Ivani A, Giubergia A, Garofalo D, Caglio A, Brazzoli E, Lipari R, Carrozza MC et al (2025) Applications of robotics for autism spectrum disorder: a scoping review. *Rev J Autism Dev Disord* 12(3):455–476. <https://doi.org/10.1007/s40489-023-00402-5>
22. Kewalramani S, Allen K-A, Leif E, Ng A (2024) A scoping review of the use of robotics technologies for supporting social-emotional learning in children with autism. *J Autism Dev Disord* 54(12):4481–4495. <https://doi.org/10.1007/s10803-023-06193-2>
23. Wang W, Xiao J, Diao L (2025) The effects of robots on children with autism spectrum disorder: a meta-analysis. *J Autism Dev Disord* 1–12. <https://doi.org/10.1007/s10803-025-06883-z>
24. Robins B, Dautenhahn K, Nadel J (2018) Kaspar, the social robot and ways it may help children with autism—an overview. *Enfance* N° 1(1):91–102. <https://doi.org/10.3917/enf2.181.0091>
25. Robins B, Dautenhahn K, Boekhorst RT, Billard A (2005) Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Univ Access Inf Soc* 4(2):105–120. <https://doi.org/10.1007/s10209-005-0116-3>
26. Pop CA, Simut R, Pinteau S, Saldien J, Rusu A, David D, Vanderfaëillie J, Lefeber D, Vanderborcht B (2013) Can the social robot probo help children with autism to identify situation-based emotions? a series of single case experiments. *Int J Hum Robot* 10(3), 1350025. <https://doi.org/10.1142/S0219843613500254>
27. Marino F, Chilà P, Sfrazzetto ST, Carrozza C, Crimi I, Failla C, Busà M, Bernava G, Tartarisco G, Vagni D, Ruta L, Pioggia G (2020) Outcomes of a robot-assisted social-emotional understanding intervention for young children with autism spectrum disorders. *J Autism Dev Disord* 50(6):1973–1987. <https://doi.org/10.1007/s10803-019-03953-x>
28. Holeva V, Nikopoulou V, Lytridis C, Bazinas C, Kechayas P, Sidropoulos G, Papadopoulou M, Kerasidou M, Karatsioras C, Geronikola N, Papakostas GA, Kaburlasos VG, Evangelidou A (2024) Effectiveness of a robot-assisted psychological intervention for children with autism spectrum disorder. *J Autism Dev Disord* 54(2):577–593. <https://doi.org/10.1007/s10803-022-05796-5>
29. Erol Barkana D, Bartl-Pokorny KD, Kose H, Landowska A, Milling M, Robins B, Schuller BW, Uluer P, Wrobel MR, Zorcec T (2024) Challenges in observing the emotions of children with autism interacting with a social robot. *Int J Soc Robot* 16(11–12):2261–2276. <https://doi.org/10.1007/s12369-024-01185-3>
30. Sani-Bozkurt S, Bozkus-Genc G (2023) Social robots for joint attention development in autism spectrum disorder: a systematic review. *Int J Disabil Dev Educ* 70(5):625–643. <https://doi.org/10.1080/1034912X.2021.1905153>
31. Landowska A, Karpus A, Zawadzka T, Robins B, Barkana DE, Kose H, Zorcec T, Cummins N (2022) Automatic emotion recognition in children with autism: a systematic literature review. *Sensors* 22(4):1649. <https://doi.org/10.3390/s22041649>
32. Li J, Davison D, Alcorn A, Williams A, Dimitrijevic SB, Petrovic S, Chevalier P, Schadenberg B, Ainger E, Pellicano L et al (2020) Non-participatory user-centered design of accessible teacher-teleoperated robot and tablets for minimally verbal autistic children. In *Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, pp 1–9. <https://doi.org/10.1145/3389189.3393738>
33. Liu C, Conn K, Sarkar N, Stone W (2008) Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Trans Robot* 24(4):883–896. <https://doi.org/10.1109/TRO.2008.2001362>
34. Kim ES, Daniell CM, Makar C, Elia J, Scassellati B, Shic F (2015) Potential clinical impact of positive affect in robot interactions for autism intervention. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp 8–13. <https://doi.org/10.1109/ACII.2015.7344544>
35. Aziz AA, Mokhsin M, Moganan FFM, Ismail A, Sakamat N, Zainol AS, Lokman AM (2018) Humanoid-robot as teaching mediator: research model in demonstrating the autistic children learning motivation based on the emotional responses. *Adv Sci Lett* 24(4):2296–2300. <https://doi.org/10.1166/asl.2018.10939>
36. Dautenhahn K, Werry I (2002) A quantitative technique for analysing robot-human interactions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol 2. pp 1132–1138. <https://doi.org/10.1109/IRDS.2002.1043883>
37. Silva S, Soares F, Costa S, Pereira AP, Moreira F (2012) Development of skills in children with asd using a robotic platform. In *2012 IEEE 2nd Portuguese Meeting in Bioengineering (ENBENG)*, pp 1–4. <https://doi.org/10.1109/ENBENG.2012.6331347>
38. Yun S-S, Choi J, Park S-K (2016) Robotic behavioral intervention to facilitate eye contact and reading emotions of children with autism spectrum disorders. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp 694–699. <https://doi.org/10.1109/ROMAN.2016.7745194>
39. Alnajjar F, Cappuccio M, Renawi A, Mubin O, Loo CK (2021) Personalized robot interventions for autistic children: an automated methodology for attention assessment. *Int J Soc Robot* 13(1):67–82. <https://doi.org/10.1007/s12369-020-00639-8>
40. Ghorbandaei Pour A, Taheri A, Alemi M, Meghdari A (2018) Human–robot facial expression reciprocal interaction platform: case studies on children with autism. *Int J Soc Robot* 10(2):179–198. <https://doi.org/10.1007/s12369-017-0461-4>
41. Wang H, Hsiao P-Y, Min B-C (2016) Examine the potential of robots to teach autistic children emotional concepts: a preliminary study. In *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, 571–580. Proceedings 8, November 1–3*. https://doi.org/10.1007/978-3-319-47437-3_56
42. Rudovic O, Lee J, Mascarell-Maricic L, Schuller BW, Picard RW (2017) Measuring engagement in robot-assisted autism therapy: a cross-cultural study. *Front Robot AI* 4(36). <https://doi.org/10.3389/frobt.2017.00036>
43. Mazzei D, Billeci L, Armato A, Lazzeri N, Cisternino A, Pioggia G, Iglizzio R, Muratori F, Ahluwalia A, De Rossi D (2010) The face of autism. In *19th International Symposium in Robot and Human Interactive Communication*, pp 791–796. <https://doi.org/10.1109/ROMAN.2010.5598683>
44. Guo R, Guo H, Wang L, Chen M, Yang D, Li B (2024) Development and application of emotion recognition technology—a systematic literature review. *BMC Psychol* 12(1):95. <https://doi.org/10.1186/s40359-024-01581-4>
45. Udahemuka G, Djouani K, Kurien AM (2024) Multimodal emotion recognition using visual, vocal and physiological signals: a review. *Appl Sci* 14(17):8071. <https://doi.org/10.3390/app14178071>
46. Ramaswamy MPA, Palaniswamy S (2024) Multimodal emotion recognition: a comprehensive review, trends, and challenges. *WIREs Data Min & Knowl* 14(6):1563. <https://doi.org/10.1002/widm.1563>
47. Hareli S, Kafetsios K, Hess U (2015) A cross-cultural study on emotion expression and the learning of social norms. *Front Psychol* 6:1501. <https://doi.org/10.3389/fpsyg.2015.01501>
48. Schiavo F, Campitiello L, Todino MD, Di Tore PA (2024) Educational robots, emotion recognition and asd: new horizon in special education. *Educ Sci* 14(3):258. <https://doi.org/10.3390/educsci14030258>
49. Boluk N, Kose H (2025) Gaze analysis of children with autism during robot-assisted therapy. In: *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. Association for Computing Machinery, New York, NY, USA, Article 123, 1–6. <https://doi.org/10.1145/3715669.3727351>

50. Bicer E, Kose H (2024) Lite-fer: a lightweight facial expression recognition framework for children in resource-limited devices. In 2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG), pp 1–9. <https://doi.org/10.1109/FG59268.2024.10581970>
51. European parliament: EU AI act: first regulation on artificial intelligence. (2023). <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. Accessed: 2026-02-06
52. European commission: AI act enters into force (2024). https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en. Accessed: 2026-02-06
53. European commission: regulatory framework for artificial intelligence (AI act). (2024). <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>. Accessed: 6 Feb 2026
54. Chen H, Ns W (2014) The assignment of scores procedure for ordinal categorical data. *ScientificWorldjournal* 2014:1–7. <https://doi.org/10.1155/2014/304213>
55. Daniel Fernández RCIL, Gu PY (2020) Assigning scores for ordered categorical responses. *J Educ Chang Appl Stat* 47(7):1261–1281. <https://doi.org/10.1080/02664763.2019.1674790>
56. Scassellati B, Admoni H, Matarić M (2012) Robots for use in autism research. *Annu Rev Biomed Eng (Volume 14(1))*:275–294. <https://doi.org/10.1146/annurev-bioeng-071811-150036>
57. Diehl JJ, Schmitt LM, Villano M, Crowell CR (2012) The clinical use of robots for individuals with autism spectrum disorders: a critical review. *Res Autism Spectr Disord* 6(1):249–262. <https://doi.org/10.1016/j.rasd.2011.05.006>
58. Adolphs R, Mlodinow L, Barrett LF (2019) What is an emotion? *Curr Biol* 29(20):R1060–R1064. <https://doi.org/10.1016/j.cub.2019.09.008>
59. Gunes H, Schuller B, Pantic M, Cowie R (2011) Emotion representation, analysis and synthesis in continuous space: a survey. In 2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG, pp 827–834. <https://doi.org/10.1109/FG.2011.5771357>
60. Ortony A, Turner (1990) What's basic about basic emotions? *Psycho. Rev* 97(3):315. <https://doi.org/10.1037/0033-295X.97.3.315>
61. Rudovic O, Lee J, Dai M, Schuller B, Picard RW (2018) Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci Robot* 3(19). <https://doi.org/10.1126/scirobotics.aao6760>
62. Gay V, Leijdekkers P, Wong F (2013) Using sensors and facial expression recognition to personalize emotion learning for autistic children. *Stud Health Technol Inf* 189:71–76. <https://doi.org/10.3233/978-1-61499-268-4-71>
63. Abdullah SMSA, Ameen SYA, Sadeeq MA, Zeebaree S (2021) Multimodal emotion recognition using deep learning. *JASTT* 2(01):73–79. <https://doi.org/10.38094/jastt20291>
64. Hirokawa M, Funahashi A, Itoh Y, Suzuki K (2014) Design of affective robot-assisted activity for children with autism spectrum disorders. In IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions, pp 365–370. <https://doi.org/10.1109/ROMAN.2014.6926280>
65. Landowska A, Zawadzka T, Zawadzki M (2022) Mining inconsistent emotion recognition results with the multidimensional model. *IEEE Access* 10:6737–6759. <https://doi.org/10.1109/ACCESS.2021.3139078>
66. Milling M, Baird A, Bartl-Pokorny KD, Liu S, Alcorn AM, Shen J, Tavassoli T, Ainger E, Pellicano E, Pantic M, Cummins N, Schuller BW (2022) Evaluating the impact of voice activity detection on speech emotion recognition for autistic children. *Front Comput Sci* 4. <https://doi.org/10.3389/fcomp.2022.837269>
67. Javed H, Jeon M, Park CH (2018) Adaptive framework for emotional engagement in child-robot interactions for autism interventions. In 2018 15th International Conference on Ubiquitous Robots, UR 2018, pp 396–400. <https://doi.org/10.1109/URAI.2018.8441775>
68. Landowska A (2019) Uncertainty in emotion recognition. *JICES* 17(3):273–291. <https://doi.org/10.1108/JICES-03-2019-0034>
69. Landowska A, Miler J (2016) Limitations of emotion recognition in software user experience evaluation context. In Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, FedCSIS, pp 1631–1640. <https://doi.org/10.15439/2016F535>
70. Milling M, Bartl-Pokorny KD, Schuller BW (2022) Investigating automatic speech emotion recognition for children with autism spectrum disorder in interactive intervention sessions with the social robot kaspar. medRxiv 2022–02. <https://doi.org/10.1101/2022.02.24.22271443>
71. Bal E, Harden E, Lamb D, Van Hecke AV, Denver JW, Porges SW (2010) Emotion recognition in children with autism spectrum disorders: relations to eye gaze and autonomic state. *J Autism Dev Disord* 40(3):358–370. <https://doi.org/10.1007/s10803-009-0884-3>
72. Fadhil TZ, Mandeel AR (2018) Live monitoring system for recognizing varied emotions of autistic children. In ICOASE 2018 - International Conference on Advanced Science and Engineering, pp 151–155. <https://doi.org/10.1109/ICOASE.2018.8548931>
73. Marinoiu E, Zanfir M, Oлару V, Sminchisescu C (2018) 3d human sensing, action and emotion recognition in robot assisted therapy of children with autism. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 2158–2167. <https://doi.org/10.1109/CVPR.2018.00230>
74. Liu C, Conn K, Sarkar N, Stone W (2007) Affect recognition in robot assisted rehabilitation of children with autism spectrum disorder. In Proceedings 2007 IEEE International Conference on Robotics and Automation, pp 1755–1760. <https://doi.org/10.1109/ROBOT.2007.363576>
75. Schuller B (2018) What affective computing reveals about autistic children's facial expressions of joy or fear. *Computer* 51(6):7–8. <https://doi.org/10.1109/MC.2018.2701647>
76. Sorensen T, Zane E, Feng T, Narayanan S, Grossman R (2019) Cross-modal coordination of face-directed gaze and emotional speech production in school-aged children and adolescents with asd. *Sci Rep* 9(1). <https://doi.org/10.1038/s41598-019-54587-z>
77. Grossard C, Dapogny A, Cohen D, Bernheim S, Juillet E, Hamel F, Hun S, Bourgeois J, Pellerin H, Serret S, Bailly K, Chaby L (2020) Children with autism spectrum disorder produce more ambiguous and less socially meaningful facial expressions: an experimental study using random forest classifiers. *Mol Autism* 11(1). <https://doi.org/10.1186/s13229-020-0312-2>
78. Belteki Z, Lumbresas R, Fico K, Haman E, Junge C (2022) 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. *IJERPH* 19(3):1469. <https://doi.org/10.3390/ijerph19031469>
79. MacFarlane H, Salem AC, Chen L, Asgari M, Fombonne E (2022) Combining voice and language features improves automated autism detection. *Autism Res* 15(7):1288–1300. <https://doi.org/10.1002/aur.2733>
80. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon

- O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, Schaik R, Sansone S-A, Schultes E et al (2016) Comment: the fair guiding principles for scientific data management and stewardship. *Sci DataSci Data* 3(1). <https://doi.org/10.1038/sdata.2016.18>
81. Tang TY (2016) Helping neuro-typical individuals to “read” the emotion of children with autism spectrum disorder: an internet-of-things approach. In *Proceedings of IDC 2016 - The 15th International Conference on Interaction Design and Children*, pp 666–671. <https://doi.org/10.1145/2930674.2936009>
 82. Di Palma S, Tonacci A, Narzisi A, Domenici C, Pioggia G, Muratori F, Billeci L (2017) Monitoring of autonomic response to sociocognitive tasks during treatment in children with autism spectrum disorders by wearable technologies: a feasibility study. *Comput Biol Med* 85:143–152. <https://doi.org/10.1016/j.compbiomed.2016.04.001>
 83. Liu C, Conn K, Sarkar N, Stone W (2008) Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder. *Int J Hum Comput Stud* 66(9):662–677. <https://doi.org/10.1016/j.ijhsc.2008.04.003>
 84. Silva V, Soares F, Esteves JS (2017) Mirroring and recognizing emotions through facial expressions for a robokind platform. In *ENBENG 2017 - 5th Portuguese Meeting on Bioengineering, Proceedings*, pp 1–4. <https://doi.org/10.1109/ENBENG.2017.7889480>
 85. Akinloye FO, Obe O, Boyinbode O (2020) Development of an affective-based e-healthcare system for autistic children. *Sci Afr* 9:e00514. <https://doi.org/10.1016/j.sciaf.2020>
 86. Krupa N, Anantharam K, Sanker M, Datta S, Sagar JV (2016) Recognition of emotions in autistic children using physiological signals. *Health Technol* 6(2):137–147. <https://doi.org/10.1007/s12553-016-0129-3>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Michał R. Wrobel received a Ph.D. in Computer Science from the Gdansk University of Technology in 2011. Since 2006 he has been working at the Faculty of Electronics, Telecommunications and Informatics, Department of Software Engineering, Gdansk University of Technology, where he is currently an Assistant Professor. He is also a member of the Emotions in HCI research group, where he researches software usability, affective computing and software management methods. His research interests include a modern approach to software development management, with a particular focus on the role of human factors in software engineering. He was project leader of the EMBOA project aimed at improving social robot intervention with children with autism using affective computing technologies.

Katrin D. Bartl-Pokorny received the M.Phil. degree in Applied Linguistics from the University of Graz, Austria, in 2010, and the Ph.D. degree in Medical Science from the Medical University of Graz, Austria, in 2019. She is currently a Senior Lecturer with the Division of Neonatology, Medical University of Graz, Austria, and a Postdoctoral Researcher with the Chair of Health Informatics, TUM University Hospital, Germany. Her research interests include digital health and socio-communicative behavior in typically developing children and children with neurodevelopmental disorders. Dr. Bartl-Pokorny is the Coordinator of the Erasmus+ project TE(A)CHADOPT.

Tatjana Zorcec is a Professor at the Medical Faculty and is an employee of the University Children’s Hospital in Skopje, Macedonia. She has expertise in diagnostics and treatment of children with autism and is appointed as National coordinator for autism. She has participated in many national and international scientific projects.

Duygun Erol Barkana is a Professor at Yeditepe University, Istanbul, Türkiye, and the director of the Robotics Research Laboratory. She works on robot-assisted rehabilitation systems, social assistive robots, human-robot interaction, and affective computing. She has received the Career Award from the Scientific and Technological Research Council of Türkiye (2009), the Outstanding Young Scientists Award from the Turkish Academy of Sciences (2018), and the Research Incentive Award from the Middle East Technical University Professor Mustafa Parlar Research and Education Foundation (2019). She has participated in several national and international projects in this field, including the FP7 project I-SUR (Intelligent Surgical Robotics), and the Erasmus+ projects EMBOA: Affective Loop in Socially Assistive Robotics as an Intervention Tool for Children with Autism, and TE(A)CHADOPT: Teaching Students How Children with Neurodevelopmental Disorders Adopt and Interact with Technologies.

Hatice Kose is a Professor at Istanbul Technical University, Turkey, coordinating the Cognitive Social Robotics research group (Affective Computing Lab) and a Visiting Professor at McGill University, Canada. She is working on Social Assistive Robots, human-robot interaction, affective computing and gamification, mostly in the domain of education, therapy and treatment of children with disabilities. She is part of several national and international projects in this field including FP6 project Robotcub, RISE project Etextweld (Welding of E-Textiles for Interactive Clothing), LUDI COST action COST Action “Play For Children With Disabilities, and Erasmus+ project EMBOA Affective loop in Socially Assistive Robotics as an intervention tool for children with autism.

Manuel Milling received his Bachelor of Science in Physics and in Computer Science from the University of Augsburg in 2014 and 2015, respectively and his Master of Science in Physics from the same university in 2018. He graduated with a PhD from the Technical University of Munich in Computer Science in 2025. His research interests include machine learning with a particular focus on the analysis of training dynamics, as well as the application of deep learning methodologies.

Ben Robins is a Visiting Senior Research Fellow at the School of Computer Science, University of Hertfordshire, U.K. His qualifications and many years of work experience lie in two disciplines, Computer Science and Dance Movement Therapy. He received his Ph.D. degree from the School of Computer Science at the University of Hertfordshire, focusing on assistive technology for children with autism, bringing together his expertise and experience in these two disciplines. His research, which started in 2002 in the AURORA Project and continued in the FP6, FP7, and H2020 European projects IROMEC, ROBOSKIN, and BabyRobot, respectively, investigates the potential use of robots as therapeutic or educational tools, encouraging basic communication, and social interaction skills in children with autism. His recent work, part of the Erasmus+ projects SMART and EMBOA, and UH’s KASPAR Project, has further investigated robot-assisted therapy and continued the development of the KASPAR robot as a therapeutic and educational tool (<http://kaspar.herts.ac.uk>). This included running long term studies with KASPAR and children with autism and other learning difficulties in collaboration with schools and medical centres internationally.

Björn W. Schuller received his diploma in 1999, his doctoral degree in 2006, and his habilitation and was entitled Adjunct Teaching Professor in 2012 all in electrical engineering and information technology from TUM in Munich/Germany. He is Full Professor of Artificial Intelligence and the Head of GLAM at Imperial College London/UK, Chair of CHI - the Chair for Health Informatics, TUM University Hospital, Munich, Germany, amongst other Professorships and Affiliations. He is a Fellow of the IEEE and Golden Core Awardee of the IEEE Computer Society, Fellow of the ACM, Fellow and President-Emeritus of the AAAC, Fellow of the BCS, Fellow of the ELLIS, Fellow of the ISCA, and Elected Full Member Sigma Xi. He (co-)authored 1,500+ publications (80,000+ citations, h-index=128).

Agnieszka Landowska is an Associate Professor with the Gdansk University of Technology, Poland. She is also a Leader of the Emotions in the Human-Computer Interactions Group (EMORG) and Head of the Software Engineering Department. She leads a project that develops mobile applications dedicated to autism therapy (Friendly Apps). She supervised research in project AFFITS (Methods and tools for affect-aware intelligent tutoring systems), AUTMON (Automated therapy monitoring for children with autism spectrum disorder), and EMBOA (Affective loop in Socially Assistive Robotics as an intervention tool for children with autism), and took part in UE COST LUDI (Play for children with disabilities) and SHELD-ON (Smart habitat for the elderly) projects. Her research is focused on making the technology more humane, including topics of human-computer and human-robot interaction, accessibility and adoption of technology, user experience, and affective computing.