



OPEN

DATA DESCRIPTOR

# Engagnition: A multi-dimensional dataset for engagement recognition of children with autism spectrum disorder

Won Kim<sup>1,2</sup>, Minwoo Seong<sup>1,2</sup>, Kyung-Joong Kim<sup>1</sup> & SeungJun Kim<sup>1</sup>✉

Engagement plays a key role in improving the cognitive and motor development of children with autism spectrum disorder (ASD). Sensing and recognizing their engagement is crucial before sustaining and improving the engagement. Engaging technologies involving interactive and multi-sensory stimuli have improved engagement and alleviated hyperactive and stereotyped behaviors. However, due to the scarcity of data on engagement recognition for children with ASD, limited access to and small pools of participants, and the prohibitive application requirements such as robots, high cost, and expertise, implementation in real world is challenging. However, serious games have the potential to overcome those drawbacks and are suitable for practical use in the field. This study proposes Engagnition, a dataset for engagement recognition of children with ASD ( $N = 57$ ) using a serious game, “Defeat the Monster,” based on enhancing recognition and classification skills. The dataset consists of physiological and behavioral responses, annotated by experts. For technical validation, we report the distributions of engagement and intervention, and the signal-to-noise ratio of physiological signals.

## Background & Summary

Children with autism spectrum disorder (ASD) often exhibit delays in motor development and deficits in motor skills, which may compromise their engagement in daily activities<sup>1,2</sup>. These challenges have the potential to impede the development of cognitive, social interaction, and communication skills, exposing children with ASD to fewer learning opportunities<sup>3,4</sup>. Therefore, sensing and recognizing the engagement status of children with ASD is an important step that lays the groundwork for further interventions and promotes their improved engagement<sup>5–9</sup>.

Engaging technologies that involve physical activity<sup>10–13</sup>, interactive stimuli<sup>14–16</sup>, extended reality<sup>12,17</sup>, robot<sup>18–21</sup> and serious games (SG)<sup>22–24</sup> are beneficial to children with ASD by improving their engagement and reducing stereotyped behaviors. While physical activity has helped improve the engagement of children with ASD, recent technological advancements have allowed it to be applied to the fields of interactive digital treatment or SG incorporating motion tracking, and robotics. Many studies have focused on robot applications, such as robot-enhanced therapy and socially assistive robotics<sup>18–21</sup>, whose benefits include increased motivation and engagement, reduced anxiety compared to human interactions, and structured and repetitive support—particularly advantageous for engaging children with ASD<sup>18</sup>. Implementing robotic systems in the field, however, is challenging; it requires considerable expertise<sup>25</sup> and high implementation costs<sup>26,27</sup>. These barriers reduce the feasibility of robot use for children with ASD in real-world settings.

Emphasis on artificial intelligence (AI) that is integrated with robots is growing, rather than solely relying on robot platforms. These AI techniques have the potential to solve challenging and complex problems in the field of robotics, particularly in real-world applications<sup>28</sup> (e.g., detection of facial expression<sup>29</sup> and emotional<sup>18</sup> and stereotyped behavior<sup>30</sup>, and safety monitoring<sup>31</sup>). Applying an AI approach to SG that can be implemented in classroom settings is promising in terms of cost-effectiveness, lower barriers to implementation without expertise, and fostering problem-solving and learning skills, which can improve engagement and independence in children with ASD<sup>22,32,33</sup>. In addition, the primary advantages of robots (i.e., multi-sensory audiovisual aids and

<sup>1</sup>Gwangju Institute of Science and Technology, School of Integrated Technology, Gwangju, 61005, South Korea.

<sup>2</sup>These authors contributed equally: Won Kim, Minwoo Seong. ✉e-mail: [seungjun@gist.ac.kr](mailto:seungjun@gist.ac.kr)

Study (year)	Participants (#)	Interaction and Tasks	Dataset Composition	Engagement annotation (annotator)	Data Availability
Jain <i>et al.</i> <sup>29</sup>	Children with ASD ( $N=7$ )	Space-themed math game with socially assistive robot	Facial expression and detection confidence value, head position, eye gaze direction, audio feature, game performance and elapsed time	Engaged/ disengaged (the first author)	Public
Javed <i>et al.</i> <sup>9</sup>	Children with ASD ( $N=5$ )	Sensory maze game with socially assistive robots	Video, audio and motion-tracking	0–5 Likert scale (3 human annotators)	On request
CultureNet <sup>19</sup>	Children with ASD ( $N=30$ )	Robot-assisted autism therapy for emotion expressions	Face images from video	Continuous scale from –1 to +1 (5 experts)	Non-public
PPA-net <sup>18</sup>	Children with ASD ( $N=35$ )	Robot-assisted autism therapy for emotion expressions	Facial expressions, head movements, body movements, pose, and gestures, audio, heart rate, electrodermal activity, body temperature	Continuous scale from –1 to +1 (5 experts)	Non-public
Chorianopoulou <i>et al.</i> <sup>92</sup>	Children with ASD ( $N=9$ )	Conversational interaction between parents and child	Acoustic and linguistic data	Engaged/ disengaged (1 expert)	Non-public
Fan <i>et al.</i> <sup>93</sup>	Adolescents with ASD ( $N=20$ )	Virtual driving	Electroencephalography, eye gaze	6-levels (1 expert)	Non-public
Engagnition* (Our dataset) <sup>53</sup>	Children with ASD ( $N=57$ )	Serious game based on physical activity (low-and-high physical exertions)	Gaze fixation, intervention, GSR, ST, performance, elapsed time, ACC, SUS, NASA-TLX (Video and audio recordings)	3-levels (3 experts & the first author)	Public

**Table 1.** Previous studies on engagement recognition in children with ASD and details of participants, interactions and tasks, dataset composition, engagement annotation (annotator), and lastly, data availability. Engagnition\* in the last line represents the details of our dataset and study position within the field. Abbreviations are explained in the following paragraph.

structured and repetitive support) can also be used in SG, and these can practically play to their strengths since applications with AI are part of their capabilities<sup>18</sup>. Thus, it is promising to promote SG incorporating AI to advance engagement recognition, “Engagnition” for further interventions.

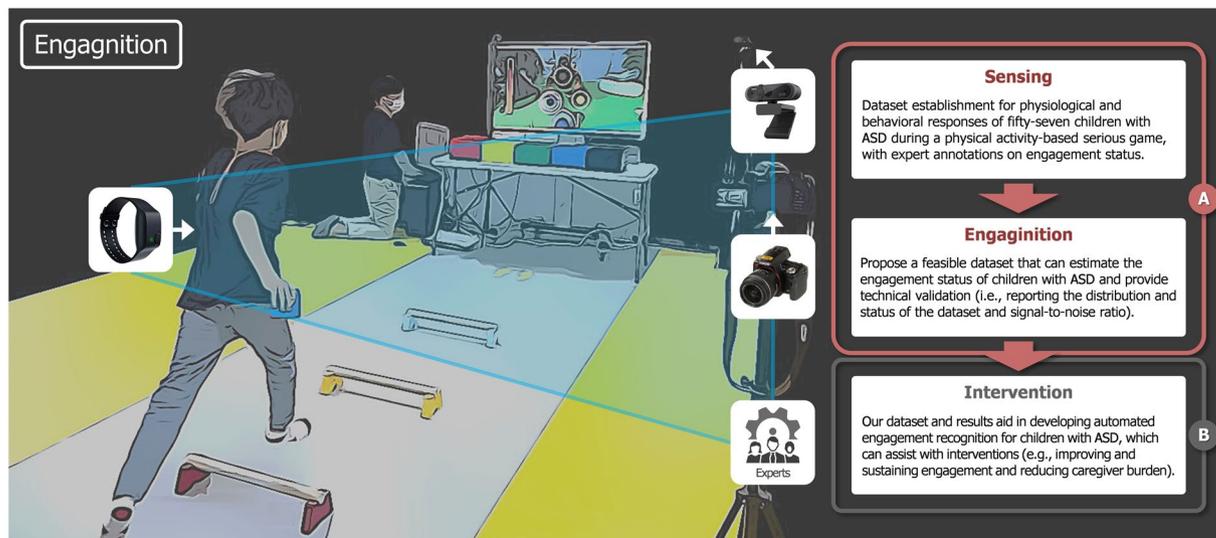
Acquiring data and recruiting children with ASD, however, can be challenging owing to limited access and small pools of participants<sup>6,34</sup>. Consequently, few datasets focused on children with ASD exist and only a limited number of datasets are available for public use. While available datasets were published in the following areas: affective computing capable of recognizing emotions or stress (i.e., AKTIVES<sup>35</sup>, ReCANVo<sup>36</sup>), social interactions between conversational partners or robots<sup>9,34</sup> (i.e., DREAM dataset<sup>37</sup>), and behavioral and gestural recognition<sup>34,37</sup> for children with ASD, datasets relating to engagement relevant to learning opportunities and functional development are scarce as shown in Table 1. Research on the engagement of children with ASD is limited mostly by non-public datasets and focuses on robot applications, which results in high barriers to practical use. Additionally, these studies rely on limited-diversity datasets.

Engagement is not a single attribute, but consists of multi-dimensional attributes in terms of the spectrum of engagement status: mental, behavioral, and emotional<sup>6</sup>. Thus, incorporating and integrating those that contain multi-dimensional sensors is crucial to ensure the Engagnition dataset. Our dataset includes expert annotations and physiological and behavioral responses: annotations on engagement<sup>38,39</sup>, gaze fixation<sup>21,40</sup>, and intervention<sup>41</sup> and responses on galvanic skin response (GSR)<sup>42–44</sup>, skin temperature (ST)<sup>45,46</sup>, performance<sup>47</sup>, elapsed time<sup>48</sup>, and accelerometer (ACC)<sup>49,50</sup>. Additionally, we included subjective questionnaires to evaluate the system usability scale (SUS)<sup>51</sup> and workload, called NASA Task Load Index (NASA-TLX)<sup>52</sup>.

Our study aims to compile the Engagnition dataset<sup>53</sup> for children with ASD, during physical activity-based SG, and evaluate the distribution of engagement and intervention, and signal-to-noise ratio (SNR) on physiological signals for technical validations. As shown in Fig. 1, our Engagnition dataset<sup>53</sup> contributes to the recognition of engagement for children with ASD during activities or SG with both static and dynamic physical activities using non-intrusive physiological and behavioral sensors with annotations by experts. Automating Engagnition with follow-up interventions may enable structured and repetitive learning from SG, improving the engagement of children with ASD and lessening the burden on special educators and parents<sup>54</sup>. In pedagogical applications for children with ASD, personalized and tailored intervention may enhance cognitive and motor development, leading to improved learning effects and skills. Lastly, gaining a better understanding of how children with ASD engage with Engagnition datasets<sup>53</sup> could potentially lead to advances in engagement recognition and engage other researchers in this critical field of study.

## Methods

**Ethics statement.** The study was approved and conducted according to the guidelines and regulations of the Institutional Review Board at Gwangju Institute of Science and Technology (approval number HR-61-04-04). The study to establish a dataset of physiological and behavioral responses, as well as subjective questionnaires for children with ASD, underwent a full review. Participants were informed about the study details, including the objectives, procedures, and the roles of parents and/or caregivers, and provided their consent to participate in the research. They agreed to be recorded during the study and consented to the use of personal data to the extent necessary for the research project.



**Fig. 1** Overview of Engagnition dataset collection for children with ASD; Apparatus in the white squares, Empatica E4 wristband, and camera for expert annotation comprise Engagnition dataset; (A) the scope of dataset establishment covered in this paper (Sensing and Engagnition), (B) application areas where our dataset and results can contribute (Intervention).

Baseline				LPE				HPE			
P	Age	Gender	Diagnosis	P	Age	Gender	Diagnosis	P	Age	Gender	Diagnosis
P01	11	M	ASD	P20	11	M	ASD	P39	11	M	ASD
P02	9	M	ASD, ID	P21	9	M	ASD	P40	9	M	ASD, ID
P03	11	F	ASD	P22	10	M	ASD	P41	11	F	ASD
P04	10	F	ASD	P23	11	F	ASD	P42	10	F	ASD
P05	10	F	ASD, ID	P24	10	F	ASD	P43	10	F	ASD, ID
P06	10	F	ASD	P25	10	F	ASD, ID	P44	10	F	ASD
P07	4	M	ASD	P26	10	F	ASD	P45	10	M	ASD, ID
P08	13	M	ASD	P27	11	M	ASD, ID	P46	10	M	ASD
P09	10	M	ASD, ID	P28	6	F	ASD, ID	P47	8	M	ASD, ID
P10	10	M	ASD	P29	13	M	ASD	P48	13	M	ASD
P11	7	M	ADHD	P30	10	M	ASD, ID	P49	16	M	ASD
P12	11	M	ASD	P31	10	M	ASD	P50	12	M	ASD
P13	12	M	ASD, ID	P32	7	M	ADHD	P51	12	M	ASD, ID
P14	8	F	ASD, ID	P33	11	M	ASD	P52	10	M	ASD
P15	10	M	ASD	P34	12	M	ASD, ID	P53	10	M	ASD
P16	9	M	ASD, ID	P35	8	F	ASD, ID	P54	10	M	ASD, ID
P17	11	M	ASD, ID	P36	9	M	ASD, ID	P55	11	M	ASD, ID
P18	12	M	ASD	P37	12	M	ASD	P56	7	M	ADHD
P19	6	F	ASD, ID	P38	4	M	ASD	P57	8	F	ASD, ID

**Table 2.** Demographics and diagnostic properties of study participants across conditions for the Engagnition dataset<sup>53</sup>; LPE: low physical exertion, HPE: high physical exertion, P: participant, M: male, F: female.

**Participants.** Fifty-seven children with ASD (age:  $M = 9.9$ ;  $SD = 2.1$ , gender ratio: male = 40; female = 17) between 4 and 16 years old were recruited from “Dream Tree Children Education Center,” an institution that supports special education within the community and promotes special education on basic motor functions, physical strength, and mental and physical health. They ranged from low- to high-functioning ASD, and some participants were diagnosed with an intellectual disability (ID) ( $N = 23$ ) and attention-deficit/hyperactivity disorder (ADHD) ( $N = 3$ ). Details on demographics and individual properties of our study participants are illustrated in Table 2.

Parents of children with ASD volunteered to help in the dataset acquisition in response to the recruitment notice at the center. On the day of participation, children with ASD were accompanied by their parents or caregivers. Then they were separated into different spaces, with assigned roles: the study participant (i.e., a child with ASD) was in the data collection testbed, while the parent or caregiver was in the observation area for subjective assessment on the SG engagement of participant. This included the questionnaires of usability



**Fig. 2** Presentation of SG in-game elements (i.e., monster and given tasks) according to the level of difficulty.

(i.e., SUS)<sup>51</sup> and workload (i.e., NASA-TLX)<sup>52</sup> assessment, which are completed by proxy users (i.e., their parent or caregiver) familiar with the study participants. Proxy users performed assessments on behalf of children with ASD for more reliable outcomes<sup>55,56</sup>. For the role in dataset acquisition, a compensation of \$75 was provided to the parent or caregiver, including a reward for the child with ASD.

**Serious game, “Defeat the Monster.”** To formulate an SG tailored for children with ASD for the Engagnition dataset, we interviewed a set of experts as an initial design step and conducted a series of iterative pilot studies<sup>57</sup>. The initial design process involved selecting the theme of the SG, designing the tasks to be performed in the game, and determining different levels of difficulty. The feedback and insights derived from expert interviews were then incorporated into a theoretical approach to develop an engaging SG<sup>23,58,59</sup>. The series of pilot studies was designed to verify the usability and applicability of the SG. The group of experts consulted in this initial step comprised 10 teachers with over 13 years of collective experience in on-site special education with students with ages between from 4 and 16 years to match the age ranges represented in the dataset.

In developing the SG, we considered that suiting the target population and aligning well with their levels of skill were crucial concerns. The target population of our study group was defined as children with ASD aged 4 years and older, which falls within age ranges ( $\geq 4$  years) for which such games are considered feasible. For example, exergames with more rigorous physical exertions such as dancing<sup>60,61</sup> have been investigated, whereas the SG described here only includes walking and classifying activities. Similarly, an SG was also investigated that required more demanding interaction with a virtual character for social communication<sup>32</sup> compared to the interaction demands of our SG (i.e., recognizing and classifying activities). Therefore, the SG implemented for the Engagnition dataset was designed for children with ASD aged between 4 and 16 years.

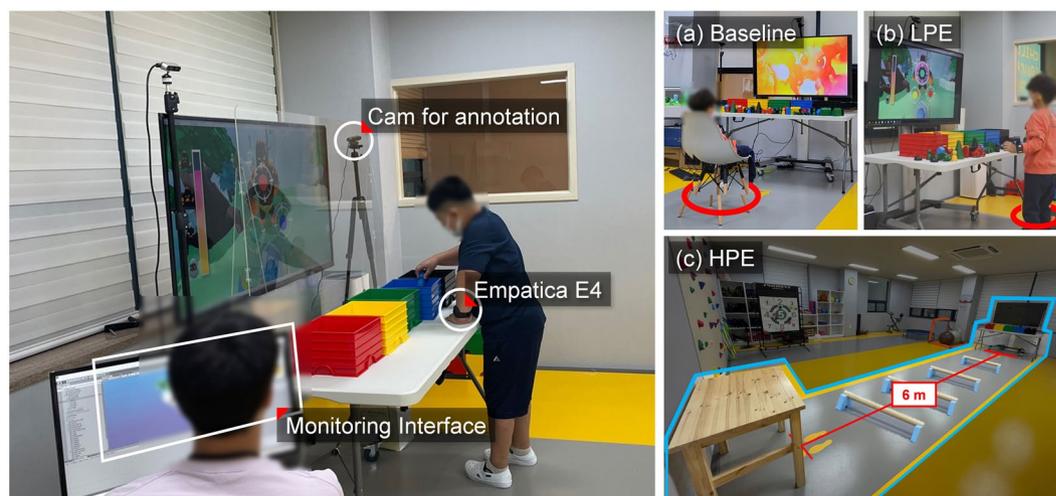
We developed “Defeat the Monster” as an SG designed to emulate the Nintendo Ring Fit. It involves two different levels of physical exertion, designated as low (LPE) and high physical exertion (HPE), which include both static and dynamic activities. Applying exergames (i.e., videogames that involve physical exertion<sup>62</sup>) in SGs has been a topic of research to support increased engagement and motivation in children with ASD<sup>23,58,63–65</sup> given their applicability and simplicity in implementation. With evident benefits in reducing repetitive behaviors as well as improving motor skills and behavior on executive function tasks<sup>64,65</sup>, exergames have been adopted in tandem with learning activity in the classroom<sup>66</sup>. Their practical applications have expanded to in-class activity and pedagogical methodology (based on a consensus of more than eight experts) using commercial games such as Nintendo’s Ring Fit Adventure<sup>62</sup>. In this study, the SG was designed to enhance cognitive and motor development by incorporating visual perception, motor planning, and execution of movements through the recognition and classification of toy blocks<sup>67,68</sup>.

The SG begins with a storyline where players defeat a monster by throwing colored balls at it. To attack the monster presented on the screen, players must recognize the highlighted effect on the monster, then pick up the same color toy block from those randomly placed on a desk in front of the screen (i.e., LPE) and far away from the screen (i.e., HPE), and classify it into a box of matching color on the desk right in front of the screen. Five colors are provided: red, yellow, green, blue, and black, and a highlighted effect is randomly assigned to each trial<sup>35,69</sup>. Every time a color is presented on the screen, a highlighted effect appears on the weak point of the monster. Each time an attack is made by classifying the toy block, the score is calculated and the anger gauge of the monster increases. The level of difficulty varies based on the acquired score, as illustrated in Fig. 2 and Table 3. When the anger gauge reaches its maximum, the SG finally ends.

The level of difficulty was elevated by increasing the number of task colors presented on the screen that the players were assigned to recognize for classification in three levels<sup>70,71</sup> and decreasing reward points<sup>70</sup> as shown in Table 3. Increasing levels of difficulty is important in designing SGs because it leads to increasing competence in particular skills for children with ASD by providing challenging yet achievable goals in given tasks<sup>58,72</sup>. Thus, defining well-mediated levels of difficulty is crucial to prevent children with ASD from becoming frustrated by drastic changes such as situations where levels of difficulty change rapidly or the type of gameplay becomes inconsistent, as one of the experts pointed out. Involving more distractors in a series of distinct stages<sup>59,73,74</sup> and increasing the speed of the game<sup>75,76</sup> are prevalent techniques to increase levels of difficulty in SGs for children with ASD. As the goal of our SG was to enhance recognition and classification skills<sup>67,68</sup>, the level of difficulty was designed to increase gradually by presenting more colors at each level from 1 to 3 by adding more visual distractors and aggravating the perceptual demands on the player<sup>59,71,77</sup>. The difficulty increases in levels based on defined ranges of cumulative scores<sup>58</sup>. Stronger-looking types of monsters appear at higher levels of difficulty;

Level of difficulty	Score range	Score (success/failure)	Monster type	Number of colors given by task
Level 1	0–20	+10/+5	Egg monster	1-color (Fig. 2(a))
Level 2	20–40	+4/+2	Hatched monster	2-colors (Fig. 2(b))
Level 3	40–60	+2/+1	Boss monster	3-colors (Fig. 2(c))

**Table 3.** Differential scoring logic and in-game elements applied to SG according to the level of difficulty.



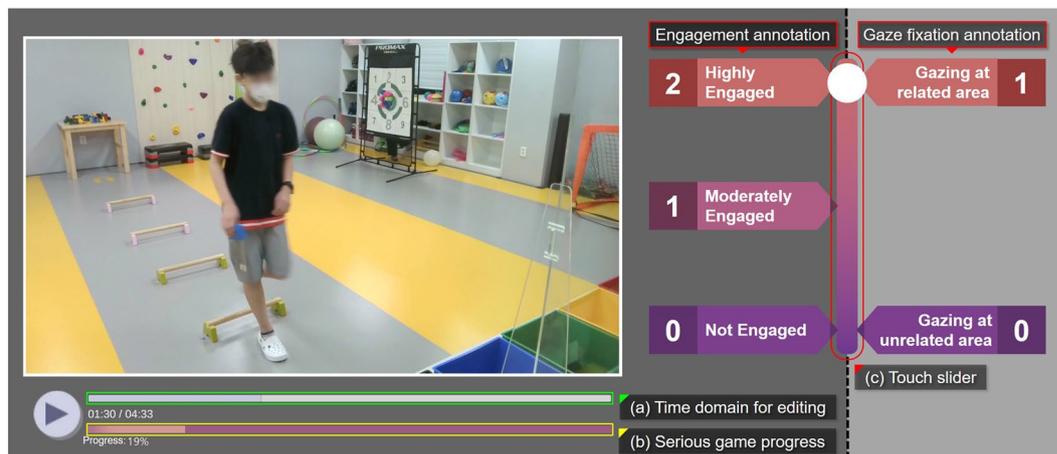
**Fig. 3** Experimental setup for dataset acquisition under the following conditions: (a) baseline, (b) LPE, and (c) HPE.

for example, an egg monster appears on difficulty level 1, a hatched monster appears on level 2, and a boss monster appears on level 3. Reward scores also decrease progressively at higher difficulty levels<sup>70</sup>. The scoring logic is designed to grant points even if the player fails (i.e., no decrease in points for failed tasks)<sup>78</sup> to prevent children from being frustrated and discouraged by failing to hit the monster's weak point. Levels of difficulty were not counterbalanced due to the hypersensitivity of children with ASD<sup>79–81</sup>.

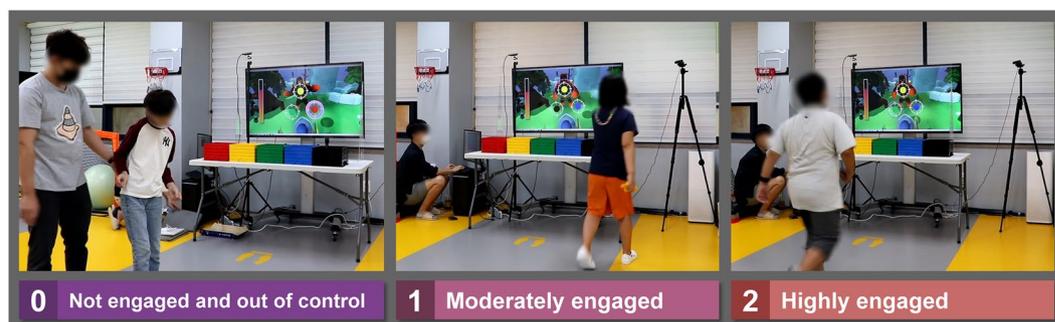
An iterative series of pilot studies was conducted to ensure the applicability of our SG for children with ASD before data were collected, initially with groups of typical participants and subsequently with experts and children with ASD. The first pilot study was conducted to identify and verify usability issues from the perspective of two typical participants (ages 20 and 23) with backgrounds in the field of human-computer interaction from undergraduate and graduate courses. A follow-up pilot study was conducted at an on-site location where data were collected. This involved three experts with over five years of experience in special education and two children with ASD at the median age of the Engagnition dataset (9 and 10) to discover any on-site difficulties and problematic issues that should be improved and verify that our SG was accessible and feasible for children with ASD.

The first pilot study identified some areas of improvement for the animation displayed when the player throws a toy block. The animation was complemented to provide visual feedback on accuracy by showing a trajectory for the animated toy<sup>58,82</sup>. This is, if an attack on the monster's weak point is successful, the toy is shown hitting the monster on the screen with a sound effect indicating success. If an attack fails, the toy is shown passing to the side of the monster with a sound effect indicating a miss. Key findings from a follow-up pilot study indicated that the size of highlighted area showing the monster's weak point and that of the interface elements showing different target colors in the original version might not be accessible and clear for children with ASD. The highlighted effect presented as the monsters' weak point was shown in white in the initial version of the game. We changed this to purple as shown in (b) and (c) in Fig. 2 to avoid overlapping with our five task colors (e.g., red, yellow, green, blue, and black), because the white color could be difficult to see and discern for children with difficulty with visual sensory cues<sup>83,84</sup>. The size of the circle showing the color was therefore increased to improve accessibility and clarity to support the children in focusing on the targeted objectives<sup>85</sup>. Thus, we confirmed that young children with ASD (4 and 6 years old), including the youngest children, had no difficulties understanding and playing the SG in the main data collection process through iterative modifications in a series of pilot studies.

**Apparatus.** *Engagnition sensors and condition setups.* Owing to the hypersensitivity of children with ASD, non-intrusive wearable wristbands and web camera sensors were used as shown in Fig. 3. The Empatica E4 wristband was attached to collect time-series data (i.e., ACC, GSR, and ST)<sup>86</sup>. These data streams were presented in real-time to experimenters for supervision and control purposes, along with a game running scene on the right, as shown in Fig. 3. Front and rear cameras were set up to record and observe the behavior of the children during SG. These recordings were later used to annotate engagement, gaze fixation, and intervention purposes.



**Fig. 4** Our own-developed annotation tool for engagement and gaze fixation for children with ASD. Our tool features the following functions: (a) time domain for editing the annotation, (b) overall progress of SG, and (c) touch slider.



**Fig. 5** Snapshot of annotated engagement of children with ASD during the SG, from left: low-to-high engagement status.

Experimental setups for data collection on three conditions, baseline, LPE, and HPE, were designed differently as shown in Fig. 3. The baseline data were collected for a minute, to achieve the state of maximum relaxation and minimize any effects from physical exertion after sensor stabilization. In the baseline, children sat in chairs and watched videos on YouTube based on their preferences or suggestions from parents and therapists<sup>87</sup> as shown in Fig. 3(a). In LPE, participants stood in place and performed SG by recognizing the color presented on the screen and classifying toy blocks into the box of the same color, as shown in Fig. 3(b). HPE demands higher physical activity from the basis of LPE (i.e. the same SG). For instance, HPE involves walking back and forth over a 6-meter area with four obstacles positioned at 1.2-meter intervals as shown in Fig. 3(c). These obstacles were additionally included by therapist, who mentioned, “This would make it more familiar, increase the intensity of physical exertion, and make it more helpful in engaging the children by including instruments already in use.”

**Engagement annotation tool.** We developed our own annotation tool for engagement and gaze fixation as shown in Fig. 4. This tool allows annotators to annotate using a touch slider, enabling immediate and linear annotations. It responds quickly to behavioral changes and retains the touch input. In addition, an editing function that allows pausing or returning to a previous point was applied for adjustments.

Engagement was annotated using a ternary code to serve as the ground truth for how children with ASD engaged in a given SG. To ensure reliable engagement annotation, we grouped the annotators into a team of four, which consisted of three experts with over five years of experience, working as therapists for the participants, as well as a researcher for annotation mediation and supervision. Figure 5 depicts a snapshot of an annotation interface of an expert, illustrating how the engagement of children with ASD was annotated during SG.

Gaze fixation was annotated using a binary code to determine whether children with ASD fixated their gaze on the relevant areas of the SG. These are the essential areas that require focused gazes for navigating and progressing during SG, which involve a screen that presents SG, colored-toy blocks, and colored boxes. Before annotation, these criteria were predefined to ensure consistency in annotation. Initially, we considered measuring gaze fixation with a bar and glasses-type sensors; however, we substituted with annotation owing to the following reasons: the limited coverage of the entire action area (bar), hypersensitivity, and concerns about introducing new equipment and its potential impact on performance (glasses).

Data category	Data/signal	Sampling rate	Range	Unit	Column names in dataset
Annotation	Engagement	60 Hz	0/1/2	Integer	Engagement
	Gaze fixation	60 Hz	0/1	Integer	Gaze
	Intervention	<i>On each occasion</i>	—	Time stamp	Time stamp and type
Physiological response	GSR	4 Hz	—	$\mu\text{S}$ (Siemens)	GSR
	ST	4 Hz	—	—	Tmp
Behavioral response	Performance	<i>Every session</i>	0/1	Score	Performance
	Elapsed time	<i>Every session</i>	—	s (second)	Elapsed time in each session
	ACC	32 Hz	—	$g \approx 9.81\text{m/s}^2$	X, Y, Z, SVM
Subjective questionnaire	SUS	<i>Post-experiment</i>	0–100	Score	Usability score
	NASA-TLX	<i>Post-experiment</i>	0–100	Score	Workload score

**Table 4.** Detailed descriptions of the Engagnition dataset<sup>53</sup> and the properties of each data or signal.

**Measures.** We established a dataset within four categories (e.g., annotation, physiological and behavioral response, and a subjective questionnaire) as shown in Table 4 and Fig. 6. Categorized data or signals are shown with the sampling rate, ranges from minimum to maximum (if possible), and the unit of each data type. Column names, which can be found in our dataset, are illustrated. These are provided in CSV format in time-series, except spent time and subjective questionnaire, which cannot be described in time-series.

Our annotations consist of three components: engagement, gaze fixation, and intervention. First, engagement, serving as a ground truth for how participants engaged in the SG was annotated with three-fold divisions<sup>38,39</sup>: 0 — not engaged and out of control, 1 — moderately engaged, 2 — highly engaged. Gaze fixation was annotated based on the gaze of participants that stayed on the SG-relevant area: 0 — gazing at areas unrelated to SG, and 1 — gazing at related areas<sup>88,89</sup>. Exceptions were made when gazing to confirm or agree with the therapist and experimenter about performing the SG, which was also classified as 1. Lastly, the time stamps of intervention required<sup>41</sup> for participants to proceed with SG were annotated with intervention types (e.g., discrete intervention with time stamps, continuous intervention for all time, and no need for intervention).

Our physiological responses included GSR<sup>42–44</sup> and ST<sup>45,46</sup>. GSR, often referred to as skin conductance, gauges the electrical resistance of the skin, which changes based on its moisture levels. GSR serves as an indicator of sympathetic arousal, measuring levels of attention and affective states, which are also associated with engagement<sup>42,43</sup> of participants during SG. ST provides an understanding of the surface thermal response of the body, which reacts according to changes in blood flow due to vascular resistance or arterial blood pressure<sup>45,46</sup>. ST is also associated with the arousal and attention state of the study participant. Therefore, ST was used to offer an additional layer of physiological data, complementing GSR.

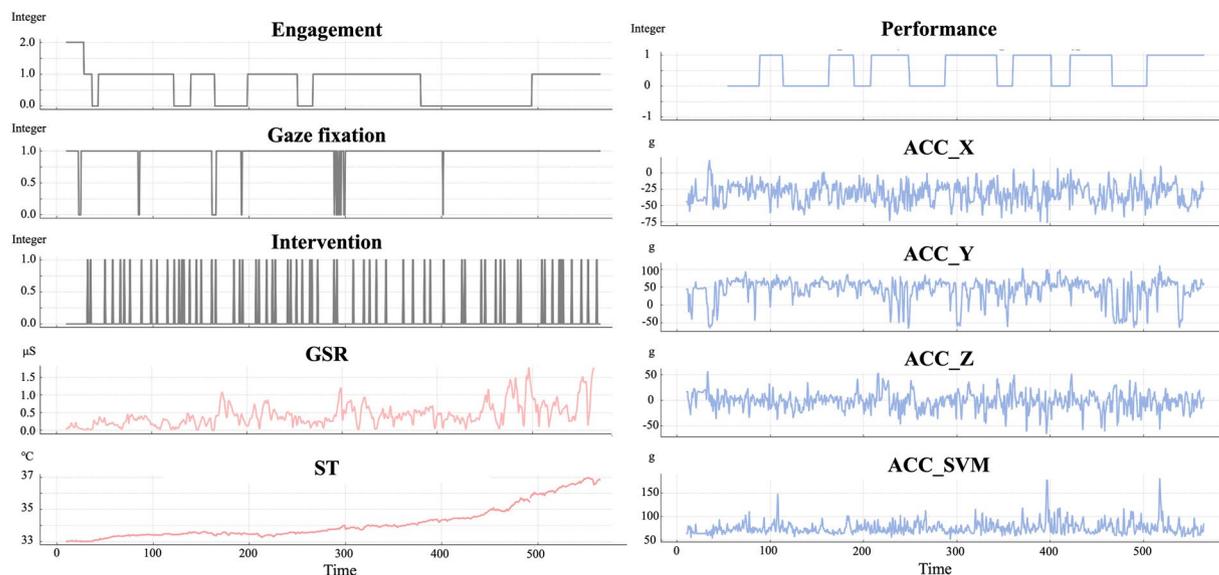
Behavioral responses included performance<sup>47</sup>, elapsed time<sup>48</sup>, and ACC. Performance, as an indicator of accuracy, denotes whether the participant successfully completed the process of recognizing the color presented and classifying the toy block into the same-colored box during each session (i.e., success/failure). This data was automatically generated from log files, and the elapsed time for each session was also recorded. ACC based on the 3-axis (x, y, z) was collected to quantify active movement, which is related to higher ACC. To capture and generalize the intensity of these movements, signal vector magnitude features were extracted from the ACC data by calculating the square root of the sum of the squared components of the vectors<sup>90,91</sup>.

Lastly, the subjective questionnaire consisted of SUS<sup>51</sup> to gauge the usability and NASA-TLX<sup>52</sup> for workload assessment. A higher score on the SUS indicates better usability, while a higher score on the NASA-TLX indicates a higher workload.

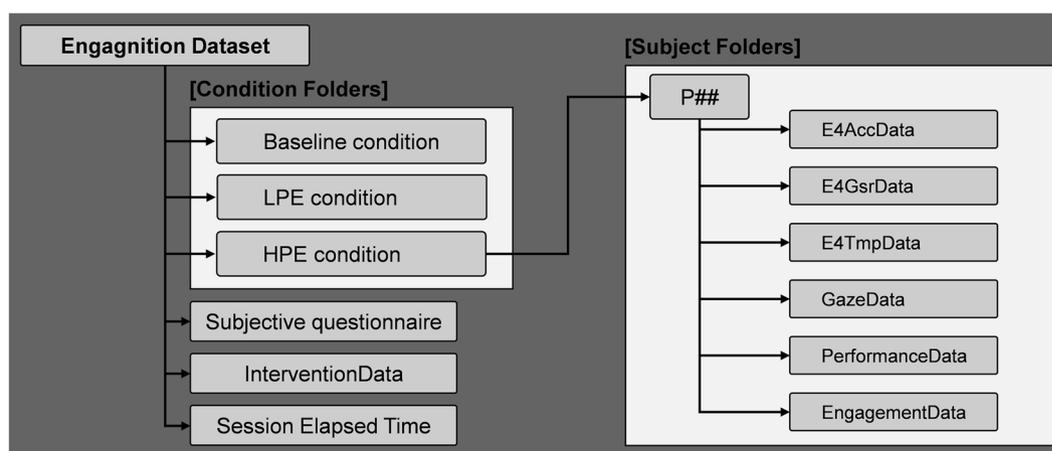
**Procedures.** Our Engagnition dataset<sup>53</sup> was established with three sequences: main data collection, post-questionnaire, and annotation. Before the main data collection, the study participants were informed about the storyline of SG, and were given a tutorial on how to navigate and complete the SG. They had practice sessions to familiarize themselves with SG and the setup for E4 sensors to their wrist, to stabilize sensor signals before the main data collection. Finally, the dataset for physiological and behavioral responses was compiled with a randomly assigned order on conditions of baseline, LPE, and HPE.

After the questionnaire, datasets for the subjective questionnaire (SUS and NASA-TLX) were compiled by caregivers and parents on behalf of the participants at the end of data collection. During the main experiment, caregivers or parents observed the SG participation of the participant through a window and a tablet PC that provided real-time mirroring. Based on those observations, they were asked to evaluate subjective questionnaires with either format of the documents (printed material) or web survey platform (SurveyMonkey).

Finally, the dataset for annotations was obtained through post-video analysis after the experiment. To ensure reliable annotations, we involved experts who have over five years of experience in the field and work as therapists for our study participants. They were familiar with the study participants and were able to better understand their intrinsic characteristics and overall performance. The therapist in charge, who participated in the data collection, primarily manipulated the touch slider, and other annotators observed together to compile, discuss, and finally modify the engagement and gaze dataset. Where any team member objected, the annotations in the data stream that caused objection were re-examined until a consensus was reached among all team members. In the case of the interventions, time stamps of intervention were generated from the actual number of interventions provided during SG and potential additional interventions required by experts.



**Fig. 6** Data stream visualization of the Engagnition dataset<sup>53</sup>. These data are samples from one of the participants; the gray graph indicates annotation data, the red graph signifies physiological data, and the blue graph denotes behavioral data.



**Fig. 7** Visual representations on the structure and folder for Engagnition dataset<sup>53</sup>.

### Data Records

The Engagnition dataset<sup>53</sup> can be accessed on figshare. We established the multi-dimensional Engagnition dataset<sup>53</sup> spanning 270.6 minutes from 57 participants, each exhibiting varying symptoms of ASD. Data were gathered under baseline, LPE, and HPE conditions. The baseline condition consisted of 64.88 minutes, averaging 3.41 minutes per participant with a standard deviation of 0.53 minutes. The LPE data spanned 69.60 minutes, with an average contribution of 3.66 minutes per participant and a standard deviation of 1.80 minutes, whereas the cumulative duration of HPE data was 136.08 minutes, with participants contributing an average of 7.16 minutes, with a standard deviation of 2.62 minutes.

All the data are stored in both CSV and XLSX formats, with the data of each participant organized within individual folders. The organization of the dataset is illustrated in Fig. 7, which provides a visual representation of the structure of the dataset. The dataset is presented with hierarchical structure, starting with the top-tier directory labeled “Engagnition Dataset.” Within this primary folder, there are three general files: a subjective questionnaire file, an interventionData file, and a file noting the elapsed time in each session. Specific folders can be used under different conditions. Within these condition-specific folders, sub-folders designated by participants identification numbers contain individual datasets, such as E4AccData, E4GsrData, E4TmpData, GazeData, PerformanceData, and EngagementData. Notably, the Baseline does not have SG participation, thus it does not involve the annotation of engagement, gaze fixation, and intervention, the behavioral responses of performance and elapsed time, and the subjective questionnaires, SUS and NASA-TLX, due to the incomparable nature between SG and the interaction of the baseline established for the maximum relaxation condition. Our dataset is accessible in the dataset repository. A concise overview of the data types available is presented

Condition	Annotation code (0)	Annotation code (1)	Annotation code (2)
LPE (%)	83,069 (28.38%)	57,899 (19.78%)	151,666 (51.82%)
HPE (%)	31,835 (7.17%)	161,814 (36.43%)	250,508 (56.4%)
Total (%)	114,904 (15.59%)	219,713 (29.82%)	402,174 (54.58%)

**Table 5.** Distributions of annotated engagement values (i.e., code 0,1, and 2) across LPE and HPE conditions: code 0 denotes participant is not engaged and out of control, code 1 indicates moderately engaged, and code 2 signifies highly engaged.

in Table 4. For more detailed specifications and descriptions of the data, please refer to the accompanying README.txt file.

**CSV file details.** The CSV file comprises Unix timestamps for the dataset and corresponding data streams. The subsequent subsections delve into the specifics of each data stream.

*E4AccData.csv.* This dataset encompasses accelerometer values for the non-dominant hand, expressed in  $m/s^2$ , across six rows. The first row denotes the SG timestamps, the second row is designated for the Unix timestamps, the third row is designated for the X values of the accelerometer, the fourth row corresponds to the Y values, the fifth row is allocated to the Z values, and the sixth row contains the signal vector magnitude values.

*E4GsrData.csv.* This dataset encompasses GSR values for the non-dominant hand, expressed in micro Siemens ( $\mu S$ ) units, across three rows. The first row denotes the SG timestamps, the second row contains the Unix timestamps and the third row contains the GSR values.

*E4TmpData.csv.* This dataset encompasses ST values for the non-dominant hand, expressed in degrees Celsius ( $^{\circ}C$ ), across three rows. The first row denotes the SG timestamps, the second row represents the Unix timestamps and the third row contains the ST values.

*GazeData.csv.* This dataset encompasses gaze fixation annotation values, expressed in 0 or 1 integers, across two rows. 0 denotes the absence of visual attention directed towards the SG-related area, whereas 1 signifies the presence of visual attention directed towards the SG-related area. The first row contains the SG timestamps, and the second row contains the gaze fixation annotation values.

*PerformanceData.csv.* This dataset encompasses performance annotation values, expressed in 0 or 1 integers, across two rows. 0 represents execution failure, while 1 signifies execution success in SG. The first row contains the SG timestamps, and the second row contains the performance values.

*EngagementData.csv.* The dataset contains engagement annotation values, which are represented by integers 0, 1, or 2, laid out across two rows. The first row contains the SG timestamps, and the second row contains the engagement values. The notation is such that 0 represents no engagement, 1 indicates moderate engagement, and 2 signifies high engagement.

**XLSX file details.** The provided XLSX file incorporates three main categories: ‘Subjective questionnaire’, which captures usability and workload by the proxy user; ‘InterventionData’, marking the time stamps when the intervention is required; and ‘Session Elapsed Time’, offering time duration spent in each session. The following subsections will delve into the specifics of each data category.

*Subjective Questionnaire.xlsx.* The ‘Subjective questionnaire’ section captures responses from both the SUS and NASA-TLX, with distinct datasets organized into separate tabs for clarity. For the NASA-TLX, columns represent the experimental ‘Condition’, ‘Participant’s identification number’, scores for individual items, and both ‘Unweighted’ and ‘Weighted’ aggregate scores. Similarly, the SUS dataset outlines the ‘Condition’, ‘Participant’s identification number’, individual query scores, and the final ‘Total Score’.

*InterventionData.xlsx.* The ‘InterventionData’ dataset has four primary rows: ‘Participant’s identification number’, ‘Condition’, ‘Intervention Type’, and ‘Timestamp of Intervention.’ Within the ‘Intervention Type’, are three categories: ‘Discrete’, ‘No Need for Intervention’, and ‘Continuous Intervention for All Time.’ The ‘Discrete’ category captures interventions that occur at specific moments, ‘No need for Intervention’ indicates that intervention is not required during the entire SG, and lastly ‘Continuous Intervention for All Time’ denotes that interventions were consistently required during the entire SG.

*Session Elapsed Time.xlsx.* The ‘Session Elapsed Time’ represents the time spent in each session of the SG, measured in seconds. The data files include ‘Condition’, ‘Participant’s identification number’, and ‘Elapsed Time per Session.’

LPE	Participant	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29
	Number	0	0	30	0	2	0	Continuous	Continuous	33	Continuous
	Participant	P30	P31	P32	P33	P34	P35	P36	P37	P38	
	Number	5	0	0	0	0	40	15	0	23	
HPE	Participant	P39	P40	P41	P42	P43	P44	P45	P46	P47	P48
	Number	0	37	7	21	2	Continuous	4	1	2	Continuous
	Participant	P49	P50	P51	P52	P53	P54	P55	P56	P57	
	Number	0	0	0	0	66	0	Continuous	9	44	

**Table 6.** The number of interventions required for each participants: The term “Continuous” signifies that participants engaged in the SG with assistance and intervention from the therapist during data collection.

Condition	Sensor	Mean	Std	Min	Q15	Q25	Q50	Q75	Q95	Max
Baseline	GSR	26.74	2.80	20.18	23.73	24.52	27.23	28.95	30.32	30.75
	ST	29.41	0.71	28.06	28.80	29.02	29.43	29.77	30.26	31.38
LPE	GSR	26.71	3.10	19.33	24.50	25.23	26.74	28.07	30.58	34.29
	ST	29.43	2.30	25.89	27.27	27.53	28.61	31.11	33.48	33.50
HPE	GSR	26.57	4.34	14.96	23.37	24.49	27.03	30.27	31.98	32.04
	ST	31.68	4.63	13.28	30.67	31.32	32.49	33.71	35.25	36.02
Total	GSR	26.67	3.48	14.96	23.59	24.64	27.03	28.85	31.34	34.29
	ST	30.17	3.20	13.28	28.15	28.83	29.91	31.87	34.72	36.02

**Table 7.** Statistics on signal-to-noise ratio (SNR) of the physiological signal acquired via Empatica E4 wristband, Q means quantiles, All the SNR values are given in dB.

### Technical Validation

To validate our Engagnition dataset<sup>53</sup>, we implemented technical validations on the distributions of annotated engagement and intervention, and the SNR of physiological signals. First, we delineated the distribution of engagement annotation values for all study participants across two distinct physical exertion conditions (i.e., LPE and HPE), as outlined in Table 5. The distribution of annotated engagement values varies between study participants. While some participants consistently exhibited an engagement annotation coded 2, others showed a more diverse distribution across codes 0, 1, and 2. Overall, code 2 engagement was the most prevalent, followed by codes 1 and 0, in that order.

In addition, we categorized the number of interventions for each participant in a tabulated format (see Table 6). Out of the 38 participants in the LPE and HPE conditions, 6 required continuous interventions throughout the sessions. Excluding these 6 participants, the frequency of interventions between participants varied widely, ranging from a minimum of 0 to a maximum of 66 interventions. In-depth temporal analysis of each intervention are in the ‘InterventionData.xlsx’ file.

Lastly, we assessed the quality of physiological signals (GSR and ST) from the Empatica E4 device using the SNR. To determine the SNR, we employed the autocorrelation function and used a second-order polynomial for a precise fit. Each set of physiological data was individually analyzed for its SNR using the decibel (dB) unit. For this estimation, we relied on the original, unaltered signals. Comprehensive statistics on SNR values are presented in Table 7. For GSR, the average SNR was 26.67 dB, with a variation of 3.48 dB. The median SNR for GSR was 27.03 dB, ranging from a low of 14.96 dB to a high of 34.29 dB. ST had an average SNR of 30.17 dB, with a variation of 3.20 dB. Its median SNR was 29.91 dB, and all values were between 13.28 dB and 36.02 dB. Additional subgroup details are available in Table 4. In line with our SNR analysis, the Empatica E4 data affirmed the high quality of the signals.

### Usage Notes

To optimize the processing of the Engagnition dataset<sup>53</sup>, we recommend the use of several Python libraries known for their effectiveness in preprocessing and feature extraction of physiological data. The Numpy library (<https://numpy.org>) plays a fundamental role in feature derivation from ACC, GSR, and ST signals. The SciPy library (<https://scipy.org>), equipped with a range of signal processing algorithms, including signal filtration, is essential for refining GSR signals. Additionally, the Pandas library (<https://pandas.pydata.org>) excels in resampling psychological signals at specific intervals. The Ledalab library focuses on GSR signals, providing both continuous and discrete decomposition analyses. These processing methodologies encompass engagement prediction, intervention time prediction, feature distillation, and the application of machine- and deep-learning algorithms.

The Engagnition dataset<sup>53</sup> is exclusively reserved for academic research purposes, in adherence to the stipulations set forth by the data contributors. Individuals or entities desiring to use the dataset must first accede to the End User License Agreement (EULA) located within the repository of the dataset. Upon executing this agreement, the signed document is forwarded to the Engagnition Research Group via [kimwon30@gm.gist.ac.kr](mailto:kimwon30@gm.gist.ac.kr), ensuring that correspondence is conducted through an official academic email address.

## Code availability

The Engnition software is available to the public through its official repository (<https://github.com/dailyminiii/Engnition>). This repository mainly includes the code for analyzing data distribution and technical validation.

Received: 16 October 2023; Accepted: 8 March 2024;

Published online: 15 March 2024

## References

- McCoy, S. M., Jakicic, J. M. & Gibbs, B. B. Comparison of obesity, physical activity, and sedentary behaviors between adolescents with autism spectrum disorders and without. *Journal of autism and developmental disorders* **46**, 2317–2326 (2016).
- Pan, C.-Y. Objectively measured physical activity between children with autism spectrum disorders and children without disabilities during inclusive recess settings in taiwan. *Journal of autism and developmental disorders* **38**, 1292–1301 (2008).
- Helminen, T. M. *et al.* Atypical physiological orienting to direct gaze in low-functioning children with autism spectrum disorder. *Autism Research* **10**, 810–820 (2017).
- Phillips, J. M. *et al.* Development of the stanford social dimensions scale: initial validation in autism spectrum disorder and in neurotypicals. *Molecular autism* **10**, 1–16 (2019).
- Sarapani, N., Morgan, L., Reinhardt, V. P., Schatschneider, C. & Wetherby, A. M. Evaluation of classroom active engagement in elementary students with autism spectrum disorder. *Journal of autism and developmental disorders* **46**, 782–796 (2016).
- Salam, H. Distinguishing engagement facets: An essential component for ai-based interactive healthcare. *arXiv preprint arXiv:2111.11138* (2021).
- Foster, M. E., Gaschler, A. & Giuliani, M. How can i help you? comparing engagement classification strategies for a robot bartender. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, 255–262 (2013).
- Leite, I., McCoy, M., Ullman, D., Salomons, N. & Scassellati, B. Comparing models of disengagement in individual and group interactions. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 99–105 (2015).
- Javed, H., Lee, W. & Park, C. H. Toward an automated measure of social engagement for children with autism spectrum disorder—a personalized computational modeling approach. *Frontiers in Robotics and AI* **43** (2020).
- Harbin, S. G., Davis, C. A., Sandall, S. & Fetting, A. The effects of physical activity on engagement in young children with autism spectrum disorder. *Early Childhood Education Journal* 1–13 (2022).
- Finkelstein, S. L., Nickel, A., Barnes, T. & Suma, E. A. Astrojumper: Designing a virtual reality exergame to motivate children with autism to exercise. In *2010 IEEE Virtual Reality Conference (VR)*, 267–268 (IEEE, 2010).
- Takahashi, I., Oki, M., Bourreau, B. & Suzuki, K. Designing interactive visual supports for children with special needs in a school setting. In *Proceedings of the 2018 Designing Interactive Systems Conference*, 265–275 (2018).
- Graham, T. N. *et al.* Design and evaluation of an exergaming system for children with autism spectrum disorder: The children's and families' perspective. *Frontiers in Virtual Reality* **3**, 817303 (2022).
- Nonnis, A. & Bryan-Kinns, N. Mazi: Tangible technologies as a channel for collaborative play. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–13 (2019).
- Soysa, A. I. & Al Mahmud, A. Tangible play and children with asd in low-resource countries: A case study. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 219–225 (2020).
- Zhiglova, Y. The interactive carpet-smart textile interface for children on autism spectrum disorder. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, 712–714 (2018).
- Malihi, M. *et al.* Evaluating the safety and usability of head-mounted virtual reality compared to monitor-displayed video for children with autism spectrum disorder. *Autism* **24**, 1924–1929 (2020).
- Rudovic, O., Lee, J., Dai, M., Schuller, B. & Picard, R. W. Personalized machine learning for robot perception of affect and engagement in autism therapy. *science robotics* **3**, 19 (2018). (2018).
- Rudovic, O. *et al.* CultureNet: A deep learning approach for engagement intensity estimation from face images of children with autism. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 339–346 (IEEE, 2018).
- Sidiropoulos, G. K. *et al.* Measuring engagement level in child-robot interaction using machine learning based data analysis. In *2020 international conference on data analytics for business and industry: way towards a sustainable economy (ICDABI)*, 1–5 (IEEE, 2020).
- Telishva, Z., Zhanatkyzy, A., Turarova, A., Rakhymbayeva, N. & Sandygulova, A. Automatic engagement recognition of children within robot-mediated autism therapy. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 471–472 (2020).
- Bhattacharya, A., Gelsomini, M., Pérez-Fuster, P., Abowd, G. D. & Rozga, A. Designing motion-based activities to engage students with autism in classroom settings. In *Proceedings of the 14th international conference on interaction design and children*, 69–78 (2015).
- Bossavit, B. & Parsons, S. “this is how i want to learn” high functioning autistic teens co-designing a serious game. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 1294–1299 (2016).
- Caro, K., Tentori, M., Martinez-Garcia, A. I. & Alvelais, M. Using the froggybobby exergame to support eye-body coordination development of children with severe autism. *International Journal of Human-Computer Studies* **105**, 12–27 (2017).
- Karademir Coşkun, T. The effectiveness of robot training in special education: a robot training model proposal for special education. *Interactive Learning Environments* **30**, 1092–1116 (2022).
- Clabaugh, C., Becerra, D., Deng, E., Ragusa, G. & Mataric, M. Month-long, in-home case study of a socially assistive robot for children with autism spectrum disorder. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 87–88 (2018).
- Dickstein-Fischer, L. *et al.* An affordable compact humanoid robot for autism spectrum disorder interventions in children. In *2011 Annual international conference of the IEEE engineering in medicine and biology society*, 5319–5322 (IEEE, 2011).
- Kunze, L., Hawes, N., Duckett, T., Hanheide, M. & Krajnck, T. Artificial intelligence for long-term robot autonomy: A survey. *IEEE Robotics and Automation Letters* **3**, 4023–4030 (2018).
- Jain, S., Thiagarajan, B., Shi, Z., Clabaugh, C. & Mataric, M. J. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Science Robotics* **5**, eaaz3791 (2020).
- Rad, N. M. & Furlanello, C. Applying deep learning to stereotypical motor movement detection in autism spectrum disorders. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 1235–1242 (IEEE, 2016).
- Cantin-Garside, K. D. *et al.* Detecting and classifying self-injurious behavior in autism spectrum disorder using machine learning techniques. *Journal of autism and developmental disorders* **50**, 4039–4052 (2020).
- Bernardini, S., Porayska-Pomsta, K. & Smith, T. J. Echoes: An intelligent serious game for fostering social communication in children with autism. *Information Sciences* **264**, 41–60 (2014).
- Bossavit, B. & Parsons, S. Outcomes for design and learning when teenagers with autism codesign a serious game: A pilot study. *Journal of Computer Assisted Learning* **34**, 293–305 (2018).
- Ivani, A. S. *et al.* A gesture recognition algorithm in a robot therapy for asd children. *Biomedical Signal Processing and Control* **74**, 103512 (2022).

35. Coşkun, B. *et al.* A physiological signal database of children with different special needs for stress recognition. *Scientific data* **10**, 382 (2023).
36. Johnson, K. T., Narain, J., Quatieri, T., Maes, P. & Picard, R. W. Recanvo: A database of real-world communicative and affective nonverbal vocalizations. *Scientific Data* **10**, 523 (2023).
37. Billing, E. *et al.* The dream dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy. *PLoS one* **15**, e0236939 (2020).
38. Hadfield, J. *et al.* A deep learning approach for multi-view engagement estimation of children in a child-robot joint attention task. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1251–1256 (IEEE, 2019).
39. Lahiri, U., Bekele, E., Dohrmann, E., Warren, Z. & Sarkar, N. Design of a virtual reality based adaptive response technology for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **21**, 55–64 (2012).
40. Putra, P., Shima, K. & Shimatani, K. Catchicken: A serious game based on the go/nogo task to estimate inattentiveness and impulsivity symptoms. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 152–157 (IEEE, 2020).
41. Sridhar, A. *et al.* Patterns of intervention utilization among school-aged children on the autism spectrum: Findings from a multi-site research consortium. *Research in autism spectrum disorders* **94**, 101950 (2022).
42. Betancourt, M. A., Dethorne, L. S., Karahalios, K. & Kim, J. G. Skin conductance as an *in situ* marker for emotional arousal in children with neurodevelopmental communication impairments: Methodological considerations and clinical implications. *ACM Transactions on Accessible Computing (TACCESS)* **9**, 1–29 (2017).
43. Brishetl, I., Ishimaru, S., Augereau, O., Kise, K. & Dengel, A. Assessing cognitive workload on printed and electronic media using eye-tracker and eda wristband. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, 1–2 (2018).
44. Hernandez, J., Riobo, I., Rozga, A., Abowd, G. D. & Picard, R. W. Using electrodermal activity to recognize ease of engagement in children during social interactions. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, 307–317 (2014).
45. Gao, N., Shao, W., Rahaman, M. S. & Salim, F. D. n-gage: Predicting in-class emotional, behavioural and cognitive engagement in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **4**, 1–26 (2020).
46. Wang, Z. *et al.* Emotionsense: An adaptive emotion recognition system based on wearable smart devices. *ACM Transactions on Computing for Healthcare* **1**, 1–17 (2020).
47. Bernardes, M., Barros, F., Simoes, M. & Castelo-Branco, M. A serious game with virtual reality for travel training with autism spectrum disorder. In *2015 international conference on virtual rehabilitation (ICVR)*, 127–128 (IEEE, 2015).
48. Hughes, D. E., Vasquez, E. & Nicsinger, E. Improving perspective taking and empathy in children with autism spectrum disorder. In *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)*, 1–5 (IEEE, 2016).
49. Li, B. *et al.* Social influences on executive functioning in autism: design of a mobile gaming platform. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13 (2018).
50. Plötz, T. *et al.* Automatic assessment of problem behavior in individuals with developmental disabilities. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, 391–400 (2012).
51. Weiss, P. L. *et al.* Usability of technology supported social competence training for children on the autism spectrum. In *2011 international conference on virtual rehabilitation*, 1–8 (IEEE, 2011).
52. Dodwell, A. & Trick, L. M. The effects of secondary tasks that involve listening and speaking on young adult drivers with traits associated with autism spectrum disorders: A pilot study with driving simulation. *Transportation research part F: traffic psychology and behaviour* **69**, 120–134 (2020).
53. Kim, W. *et al.* Engagnition: A multi-dimensional dataset for engagement recognition of children with autism spectrum disorder. *figshare* <https://doi.org/10.6084/m9.figshare.c.6879628.v1> (2024).
54. Wu, C. *et al.* Machine learning based autism spectrum disorder detection from videos. In *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*, 1–6 (IEEE, 2021).
55. Boyd, L. E. *et al.* Vrsocial: Toward immersive therapeutic vr systems for children with autism. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12 (2018).
56. Sofian, N. M., Hashim, A. S. & Sarlan, A. Multimedia elements in designing mobile app interface for autistic children: proxy user perspective. In *2021 International Conference on Computer & Information Sciences (ICCOINS)*, 96–101 (IEEE, 2021).
57. Hassan, A., Pinkwart, N. & Shafi, M. Serious games to improve social and emotional intelligence in children with autism. *Entertainment computing* **38**, 100417 (2021).
58. Whyte, E. M., Smyth, J. M. & Scherf, K. S. Designing serious game interventions for individuals with autism. *Journal of autism and developmental disorders* **45**, 3820–3831 (2015).
59. Tang, J. S., Chen, N. T., Falkmer, M., Bölte, S. & Girdler, S. A systematic review and meta-analysis of social emotional computer based interventions for autistic individuals using the serious game framework. *Research in Autism Spectrum Disorders* **66**, 101412 (2019).
60. Gao, Z. *et al.* Home-based exergaming on preschoolers' energy expenditure, cardiovascular fitness, body mass index and cognitive flexibility: a randomized controlled trial. *Journal of clinical medicine* **8**, 1745 (2019).
61. Liu, Z.-M., Chen, C.-Q., Fan, X.-L., Lin, C.-C. & Ye, X.-D. Usability and effects of a combined physical and cognitive intervention based on active video games for preschool children. *International Journal of Environmental Research and Public Health* **19**, 7420 (2022).
62. Caro, K., Martinez-Garcia, A. I. & Kurniawan, S. A performance comparison between exergames designed for individuals with autism spectrum disorder and commercially-available exergames. *Multimedia Tools and Applications* **79**, 33623–33655 (2020).
63. Getchell, N., Miccinello, D., Blom, M., Morris, L. & Szaroleta, M. Comparing energy expenditure in adolescents with and without autism while playing nintendo® wii™ games. *GAMES FOR HEALTH: Research, Development, and Clinical Applications* **1**, 58–61 (2012).
64. Dickinson, K. *et al.* A randomised control trial of the impact of a computer-based activity programme upon the fitness of children with autism. *Autism Research and Treatment* **2014** (2014).
65. Hilton, C. L. *et al.* Effects of exergaming on executive function and motor skills in children with autism spectrum disorder: A pilot study. *The American Journal of Occupational Therapy* **68**, 57–65 (2014).
66. Anderson-Hanley, C., Tureck, K. & Schneiderman, R. L. Autism and exergaming: effects on repetitive behaviors and cognition. *Psychology research and behavior management* 129–137 (2011).
67. Mieres, A. C., Kirby, R. S., Armstrong, K. H., Murphy, T. K. & Grossman, L. Autism spectrum disorder: an emerging opportunity for physical therapy. *Pediatric physical therapy* **24**, 31–37 (2012).
68. Dawson, G. & Watling, R. Interventions to facilitate auditory, visual, and motor integration in autism: A review of the evidence. *Journal of autism and developmental disorders* **30**, 415–421 (2000).
69. Saunders, J. A. & Knill, D. C. Visual feedback control of hand movements. *Journal of Neuroscience* **24**, 3223–3234 (2004).
70. Kim, B. *et al.* Puzzlewalk: A theory-driven iterative design inquiry of a mobile game for promoting physical activity in adults with autism spectrum disorder. *PLoS One* **15**, e0237966 (2020).
71. Kosmas, P., Ioannou, A. & Retalis, S. Moving bodies to moving minds: A study of the use of motion-based games in special education. *TechTrends* **62**, 594–601 (2018).

72. Przybylski, A. K., Rigby, C. S. & Ryan, R. M. A motivational model of video game engagement. *Review of general psychology* **14**, 154–166 (2010).
73. Serret, S. *et al.* Facing the challenge of teaching emotions to individuals with low-and high-functioning autism using a new serious game: a pilot study. *Molecular autism* **5**, 1–17 (2014).
74. Fletcher-Watson, S. *et al.* A trial of an iPad<sup>®</sup> intervention targeting social communication skills in children with autism. *Autism* **20**, 771–782 (2016).
75. Russo-Ponsaran, N. M., Evans-Smith, B., Johnson, J. K. & McKown, C. A pilot study assessing the feasibility of a facial emotion training paradigm for school-age children with autism spectrum disorders. *Journal of Mental Health Research in Intellectual Disabilities* **7**, 169–190 (2014).
76. Russo-Ponsaran, N. M., Evans-Smith, B., Johnson, J., Russo, J. & McKown, C. Efficacy of a facial emotion training program for children and adolescents with autism spectrum disorders. *Journal of Nonverbal Behavior* **40**, 13–38 (2016).
77. Vallefucio, E., Bravaccio, C., Gison, G., Pecchia, L. & Pepino, A. Personalized training via serious game to improve daily living skills in pediatric patients with autism spectrum disorder. *IEEE Journal of Biomedical and Health Informatics* **26**, 3312–3322 (2022).
78. Lepper, M. R. & Chabay, R. W. Intrinsic motivation and instruction: conflicting views on the role of motivational processes in computer-based education. *Educational Psychologist* **20**, 217–230, [https://doi.org/10.1207/s15326985ep2004\\_6](https://doi.org/10.1207/s15326985ep2004_6) (1985).
79. Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T. & Chakrabarti, B. Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 1377–1383 (2009).
80. Alcaniz Raya, M. *et al.* Machine learning and virtual reality on body movements' behaviors to classify children with autism spectrum disorder. *Journal of clinical medicine* **9**, 1260 (2020).
81. Tomchek, S. D., Huebner, R. A. & Dunn, W. Patterns of sensory processing in children with an autism spectrum disorder. *Research in Autism Spectrum Disorders* **8**, 1214–1224 (2014).
82. Whalen, C. *et al.* Efficacy of teachtown: Basics computer-assisted intervention for the intensive comprehensive autism program in los angeles unified school district. *Autism* **14**, 179–197 (2010).
83. Mazurek, M. O., Engelhardt, C. R. & Clark, K. E. Video games from the perspective of adults with autism spectrum disorder. *Computers in Human Behavior* **51**, 122–130 (2015).
84. Rosenthal, M. *et al.* Impairments in real-world executive function increase from childhood to adolescence in autism spectrum disorders. *Neuropsychology* **27**, 13 (2013).
85. Jaramillo-Alcázar, A., Arias, J., Albornoz, I., Alvarado, A. & Luján-Mora, S. Method for the development of accessible mobile serious games for children with autism spectrum disorder. *International Journal of Environmental Research and Public Health* **19**, 3844 (2022).
86. Milstein, N. & Gordon, I. Validating measures of electrodermal activity and heart rate variability derived from the ematica e4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience* **14**, 148 (2020).
87. Kushki, A., Khan, A., Brian, J. & Anagnostou, E. A kalman filtering framework for physiological detection of anxiety-related arousal in children with autism spectrum disorder. *IEEE Transactions on Biomedical Engineering* **62**, 990–1000 (2014).
88. Venker, C. E. *et al.* Comparing automatic eye tracking and manual gaze coding methods in young children with autism spectrum disorder. *Autism Research* **13**, 271–283 (2020).
89. Higuchi, K. *et al.* Visualizing gaze direction to support video coding of social attention for children with autism spectrum disorder. In *23rd International Conference on Intelligent User Interfaces*, 571–582 (2018).
90. Ward, J. A., Richardson, D., Orgs, G., Hunter, K. & Hamilton, A. Sensing interpersonal synchrony between actors and autistic children in theatre using wrist-worn accelerometers. In *Proceedings of the 2018 ACM international symposium on wearable computers*, 148–155 (2018).
91. Bayat, A., Pomplun, M. & Tran, D. A. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science* **34**, 450–457 (2014).
92. Chorianopoulou, A. *et al.* Engagement detection for children with autism spectrum disorder. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5055–5059 (IEEE, 2017).
93. Fan, J., Wade, J. W., Key, A. P., Warren, Z. E. & Sarkar, N. Eeg-based affect and workload recognition in a virtual driving environment for asd intervention. *IEEE Transactions on Biomedical Engineering* **65**, 43–51 (2017).

## Acknowledgements

This work was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1A4A1030075) and the GIST-MIT Research Collaboration grant funded by the GIST in 2024. We would like to express sincere appreciations to Director Choongman Lee of Dream Tree Children Education Center, as well as Teacher Sion Song and Geunseo Kim, for their cooperation and assistance in establishing the Engagnition dataset.

## Author contributions

W.K.: Conceptualization; Methodology; Data collection; Data Analysis; Investigation; Writing - Original Draft & Review & Editing; Visualization; Project administration. M.W.S.: Conceptualization; Data collection; Data Analysis; Technical Validation; Writing - Original Draft & Review & Editing; Visualization; Editing. K.J.K.: Writing - Review & Editing; Supervision; Funding acquisition. S.J.K.: Conceptualization; Methodology; Writing - Review & Editing; Supervision; Funding acquisition.

## Competing interests

The authors declare no competing interests. The serious game “Defeat the Monster” and annotation tool were developed for the purpose of establishing the Engagnition dataset for this study.

## Additional information

**Correspondence** and requests for materials should be addressed to S.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2024