Developing an AI-based Explainable Expert Support System for Art Therapy

JIWON KIM*, JIWON KANG*, MIGYEONG YANG*, and CHAEHEE PARK*, Sungkyunkwan University, Republic of Korea
TAEEUN KIM, CHA University, Republic of Korea
HAYEON SONG, Sungkyunkwan University, Republic of Korea
JINYOUNG HAN†, Sungkyunkwan University, Republic of Korea

Sketch-based drawing assessments in art therapy are widely used to understand individuals' cognitive and psychological states, such as cognitive impairments or mental disorders. Along with self-reported measures based on questionnaires, psychological drawing assessments can augment information regarding an individual's psychological state. Interpreting drawing assessments demands significant time and effort, particularly for large groups such as schools or companies, and relies on the expertise of art therapists. To address this issue, we propose an artificial intelligence (AI)-based expert support system called *AlphaDAPR* to support art therapists and psychologists in conducting large-scale automatic drawing assessments. In Study 1, we first investigated user experience in *AlphaDAPR*. Through surveys involving 64 art therapists, we observed a substantial willingness (64.06% of participants) in using the proposed system. Structural equation modeling highlighted the pivotal role of explainable AI in the interface design, affecting perceived usefulness, trust, satisfaction, and intention to use. However, our interviews unveiled a nuanced perspective: while many art therapists showed a strong inclination to use the proposed system, they also voiced concerns about potential AI limitations and risks. Since most concerns arose from insufficient trust, which was the focal point of our attention, we conducted Study 2 with the aim of enhancing trust. Study 2 delved deeper into the necessity of clear communication regarding the division of roles between AI and users for elevating trust. Through experimentation with another 26 art therapists, we demonstrated that clear communication enhances users' trust in our system. Our work not only highlights the potential of *AlphaDAPR* to streamline drawing assessments but also underscores broader implications for human-AI collaboration in psychological domains. By addressing concerns and optimizing communication, we pave the way for a symbiotic relationship between AI and human expertise, ultimately enhancing the efficacy and accessibility of psychological assessment tools.

CCS Concepts: • **Applied computing → Arts and humanities**; **Psychology**; • **Information systems → Expert systems**; • **Human-centered computing**; • **Computing methodologies → Artificial intelligence**;

Additional Key Words and Phrases: Art therapy, Drawing assessments, Expert support system, Explainable AI, AI, Draw-A-Person-in-the-Rain, Trust

*These authors contributed equally as the first authors.
†Corresponding author.

Authors' addresses: Jiwon Kim, jjeong416@g.skku.edu; Jiwon Kang, jiwonkang@skku.edu; Migyeong Yang, mgyang@g.skku.edu; Chaehee Park, chaeheepark@g.skku.edu, Sungkyunkwan University, Republic of Korea; Taeeun Kim, blessing75@cha.ac.kr, CHA University, Republic of Korea; Hayeon Song, songhy@skku.edu, Sungkyunkwan University, Republic of Korea; Jinyoung Han, jinyonghan@skku.edu, Sungkyunkwan University, Republic of Korea.

ACM 2160-6463/2024/8-ART
https://doi.org/10.1145/3689649

1 INTRODUCTION

Sketches provide insight for understanding an individual's psychological and cognitive state. For example, analyzing a human sketch can help identify an individual's mental state, such as stress and burnout [32, 60], mental development [48], child sexual abuse [3, 5, 30], and depression [15]. Additionally, drawing-based cognitive assessments can be used to detect symptoms of dementia [10, 24]. Hence, drawing assessments such as Draw-a-Person, Draw-A-Person-in-the-Rain, and House-Tree-Person assessments are widely used as supportive tools to initially assess the psychological state of an individual who potentially requires further examination in the research and education fields [36, 48] and clinical environments [32, 60].

Along with self-report measures based on questionnaires, drawing assessments can augment information regarding an individual's psychological state and reflect preconscious or unconscious material that verbalized expressions cannot access [2, 17, 39]. Because a sketch is a nonverbal communication tool, those who may have difficulty reading and writing, such as children and older adults, can easily participate in drawing assessments. The intuitive nature of sketches allows drawing assessments to be applied to large-scale groups such as schools [9, 14, 22] and workplaces [32]. In analyzing a drawing, experts such as art therapists or psychologists use predefined scoring scales to identify psychological indicators. These scales evaluate how a participant depicts a human figure and its surroundings (e.g., sky, rain) in a sketch. However, such manual scoring requires considerable time and cost, particularly when an assessment is conducted on numerous participants. In addition, the analysis results often depend on the experience of experts, which requires further inter-agreement among multiple experts [60].

A method that can automatically analyze drawing assessments can address the aforementioned problems and support human experts by reducing time and cost. In addition, the automatic method can standardize the analysis process for drawing assessments, which can mitigate the bias of human scorers. The following three system components should be investigated and developed to provide expert support systems for drawing assessments. First, a drawing result or sketch should be accurately analyzed; for example, whether there is a human in a sketch or the count of humans appearing in a sketch. Second, based on the analyzed results of a sketch, a score that reflects the psychological state should be automatically calculated by following art therapy practices. Third, the final analysis results should be provided to human experts with detailed explanations of the automatically generated results and useful information that can be used for further in-depth assessments such as the number of strokes and drawing sequences.

We propose an expert support system referred to as *AlphaDAPR* to assist art therapists and psychologists in conducting large-scale automatic drawing assessments. Among the widely used drawing assessments, we selected the Draw-A-Person-in-the-Rain (DAPR) assessment, which is an enhanced version of the Draw-A-Person (DAP) assessment. It provides findings available from the DAP assessment, as well as further valuable insights into participants' stress and coping mechanisms [36, 60]. The proposed system consists of three components: (i) an object detection model in a sketch, (ii) a sketch scoring model, and (iii) a dashboard that presents the obtained results to experts. First, we developed an object detection model based on YOLO-v5 [12] to analyze sketches for DAPR assessments, which showed higher performance than other well-known sketch detection methods such as Mask-R-CNN [18]. Based on information about the identified objects in a sketch, including the category (e.g., person, rain) of an object, and its location and size, the sketch scoring model calculates the score of a given drawing assessment, which reflects a participant's stress and his/her coping mechanism.

Explainability is essential to increase the intention to use an expert support system, which helps increase trust in the system [50]. Trust is one of the important factors in using expert support systems [7, 52]. Providing evidence for artificial intelligence (AI) predictions with visual aids [1] and/or textual information [45] can increase trust in the system [53]. Based on the insights gained from prior work on explainability in an AI-based system, we developed a dashboard in *AlphaDAPR* that includes not only the analysis results of a sketch and its score, but also visual, textual, and numerical evidence that explains the generation of the given results. In addition, to improve the understandability of the input sketches, we provide further information that can be useful for further in-depth assessments by art therapists or psychologists, which will help them make accurate assessments [57].

To validate user experiences in *AlphaDAPR* with a focus on how explainability and trust can be linked to intention to use the proposed system, we first applied a user-centric evaluation framework [31] in Study 1, which is a widely used evaluation framework for decision support systems to evaluate the proposed expert support system. Thus far, we have recruited 64 art therapists who are experts in DAPR assessments to conduct a survey and interview. The evaluation revealed that over 64% of the experts expressed an intention to use the proposed system, highlighting its potential to assist art therapists in managing large groups, such as school students. The results of structural equation modeling highlight the importance of explainable AI embedded in interface design to affect perceived usefulness, trust, satisfaction, and intention to use. The interview study indicated that most art therapists displayed a strong inclination to use the proposed system. However, amidst the positive responses to using *AlphaDAPR*, concerns regarding AI limitations and risks emerged, particularly centered on trust. As we determined trust as a critical determinant in *AlphaDAPR*, Study 2 was initiated to fortify this aspect. Building upon the hypothesis that clear communication regarding the roles of both AI and users fosters trust, we conducted a user study with an additional 26 art therapists through a survey. We confirmed that trust indeed heightens through clear communication, emphasizing that AI only serves as an initial support tool and that humans make final decisions and complete evaluations. These results have important implications on the widespread use of *AlphaDAPR*, highlighting the importance of clarifying the role of the proposed expert system as a supporter of art therapists, rather than replacing them. By conducting Study 1 with a focus on explainability and trust, and addressing the identified risks in AI-human communications through Study 2, we foster a synergistic partnership between AI and human expertise. This endeavor not only enhances the effectiveness and accessibility of psychological assessment tools but also sets a precedent for harmonious interaction between AI and human practitioners in the field.

2 BACKGROUND & RELATED WORK

2.1 Drawing-A-Person-in-the-Rain (DAPR) Assessment

Drawing assessments can be used to analyze patient drawings, which can capture their personalities, mental health, and psychological states [14, 22, 36, 60]. In drawing assessments, a client is asked to draw a freehand sketch with paper and a pencil, and then a clinician or art therapist investigates the given drawing based on a specified set of analytical scoring scales [13]. There are several popular drawing assessments [17] such as draw-a-person, house-tree-person, animal-drawing-story, draw-a-family, and draw-a-person-in-the-rain. Among these, the DAPR assessment provides valuable insights into participants' stress and coping mechanisms [22, 32, 36, 60].

The DAPR assessment guides a participant to *"draw a person(s) in the rain,"* where 'rain' implies stressful events or environments around the participants. Verinis et al. [56] empirically studied the DAPR assessment and categorized objects that appeared in the drawing: stress-related and resource-related objects. Stress-related objects (e.g., rain, clouds) indicate the perceived level of stress, and resource-related objects (e.g., an umbrella or a raincoat) can protect oneself against rain, implying the ability to cope with stress. To standardize the interpretation of the drawing results, Lack [34] developed a scale for DAPR assessment: 16 items for stress-related attributes and 19 items for resource-related attributes, which allowed art therapists to quantitatively assess the stress and coping behavior of the participants through the stress and resource scores, respectively. The final DAPR score was calculated by subtracting the stress score from the resource score. These scores aid in evaluating the participants' overall mental health issues [56].

2.2 AI Technology for Art Therapy

With recent advancements in AI technology and increasing interest in art therapy, there have been attempts to develop AI-based methods for art therapy [10, 48, 51]. Chen et al. [10] proposed a clock-drawing scoring on a scale from 1 to 6, based on the modern deep learning architecture (e.g., VGG-16 [54]), to estimate the severity level of dementia, a representative cognitive disorder. In addition, to classify children's mental development states, Rakhmanov et al. [48] developed a Draw-A-Person assessment dataset drawn from elementary school students, and tested traditional machine learning-and convolutional neural network-based architectures. Seo et al. [51] showed the potential benefit of using an AI-based assistant tool in their pilot study.

These studies have revealed the possibility of assessing hand drawings using AI technology. Our work goes one step further by developing an expert support system that can support art therapists and psychologists in conducting large-scale automatic drawing assessments (particularly for DAPR). The proposed expert support system includes (i) an automatic sketch analysis method, (ii) a DAPR scoring function, and (iii) an explainable dashboard to support experts effectively.

2.3 Trust and Explainability in AI-based Expert Support Systems

Increasing the intention to use, which affects the actual usage behavior [55], is important for designing an AI-based expert support system. Trust in the system is an influential factor in the intention to use, specifically in clinical fields, including mental and physical health, which significantly affects people. For example, Baldauf et al. [7] proposed a mobile self-diagnosis system and found that trust in AI performance influences overall system usage intention. Shibl et al. [52] revealed that experts were not willing to use the system if the system was deemed untrustworthy, which signifies that trust is important in an expert support system. To increase trust in an AI-based expert support system, clarifying the roles of AI and a user is crucial [35, 44]. Liao and Sundar [35] showed an increase in trust levels when the role of AI was clearly articulated, as opposed to situations where it remained ambiguous. Similarly, Nazaretsky et al. [44] suggested that affording users autonomy and control over modifying AI-generated outcomes helps alleviate concerns regarding potential AI replacement.

Explainable AI is another crucial factor for boosting the inclination to utilize an expert support system, thereby bolstering trust in the system [50]. By providing relevant explanations regarding the evidence for AI prediction and the underlying mechanism, an AI system can build trust with users. Panigutti et al. [45] proposed a system that suggests diagnostics for clinicians to discriminate a disease and showed that providing a relevant degree and textual information with the diagnostics can increase trust. Shin [53] found that explaining the rationale behind the AI model's predictions enhances trust in a movie recommendation decision support system. Similarly, the explainability of AI-based expert support systems can enhance trust in art therapy. Seo et al. [51] presented an AI-based assistant tool to analyze the House-Tree-Person assessment drawings in an online art therapy session based on the Wizard of OZ and qualitatively evaluated their tool with 10 art therapists. In their interview study, they suggested the importance of providing a sufficient explanation of AI results when designing tools for drawing assessments.

Inspired by Seo et al. [51], we developed an AI-based expert support system for drawing assessments (DAPR in particular) by focusing on explainability, which can contribute to experts' intention to use. In particular, we developed the proposed system to effectively visualize information provided by AI-based scoring models, as well as diverse information (e.g., sketch replay, participant information) regarding the sketches, to provide an understanding of the results of the drawing assessments (refer to Section 3.3 for more information). Using the system, we conducted an experiment to determine if clear communication enhances trust in Study 2. By placing emphasis on explainability and trust, we developed and evaluated the system, which were identified as crucial determinants influencing the intent to use the system.

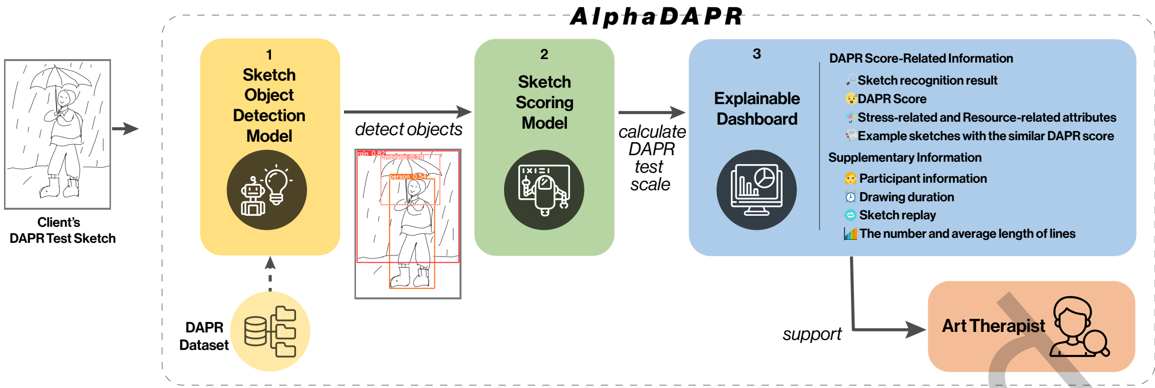

Fig. 1. Overall system process of *AlphaDAPR*

Model	Train Data	Train Size	Test Size	AP	AP_{50}	AP_{75}
	Synthesized	3,330	132	14.79	37.74	8.99
Faster R-CNN [49]	Collected	198	132	38.76	62.13	41.86
	All	3,498	132	39.54	66.96	40.99
	Synthesized	3,330	132	16.46	38.84	11.18
Mask R-CNN [18]	Collected	198	132	34.95	60.33	35.29
	All	3,498	132	37.92	65.47	37.97
	Synthesized	3,330	132	33.19	53.58	35.52
yolo-v5 [12]	Collected	198	132	44.95	67.94	51.50
	All	3,498	132	**50.46**	**74.46**	**54.68**

Table 1. Mean average precision (AP) of the object detection task on the test sets across the six object categories: person, rain, umbrella, puddle, lightning, and cloud. The baseline models are Faster R-CNN [49], Mask R-CNN [18], and yolo-v5 [12].

3 ALPHADAPR: AN AI-BASED EXPLAINABLE EXPERT SUPPORT SYSTEM FOR DAPR

In this study, we propose *AlphaDAPR*, an expert support system for art therapists and psychologists, for conducting large-scale automatic drawing assessments. As illlustrated in Figure 1, *AlphaDAPR* consists of three components: (i) a sketch object detection model, (ii) a sketch scoring model, and (iii) an explainable dashboard.

3.1 A Sketch Object Detection Model

3.1.1 DAPR Dataset. We utilized SceneDAPR [26][1], which is a scene-level sketch dataset for the DAPR assessment. Among the dataset, we used 330 scene sketches drawn by humans and 3,300 synthesized sketches, with six categories: person, rain, umbrella, lighting, puddle, and cloud, which were considered the main objects in the DAPR assessment.

3.1.2 A Scene-level Object Detection. DAPR assessment is a method used to measure objects related to stress and related resources, such as raindrops or umbrellas, and then assign scores [60]. Hence, to analyze a sketch in a DAPR assessment, we first developed a model to identify the objects drawn in a sketch. The output of the model

[1]https://github.com/DSAIL-SKKU/SceneDAPR

consisted of the coordinates of the bounding box of the identified object with a confidence score and its category information. Therefore, information regarding the objects in the sketch, their size, and location can be obtained.

For accurate sketch object detection, we considered the following representative models: Faster R-CNN [49], Mask R-CNN [18], and YOLO-v5 [12]. For evaluation, we use mean average precision (AP) with an IoU threshold of 0.5 (AP_{50}), 0.75 (AP_{75}), and from 0.5 to 0.95 (AP). We conducted early stopping with patient 50, where the training was stopped if the validation AP did not increase for 50 consecutive epochs out of the maximum 300 epochs. The performance of the models on the test set is summarized in Table 1. Among the models, we chose YOLO-v5 which outperforms the others in all APs as the sketch object detection model in the proposed system.

3.2 A Sketch Scoring Model

The sketch-scoring model takes the output of the object-detection model and calculates the DAPR assessment scale for a given sketch. *AlphaDAPR* utilizes the well-known DAPR assessment scale [25, 34]. To calculate the DAPR assessment scale, we derived the following three types of input drawing information using the output of the object detection model: (1) the stress-related score and (2) the coping resource score.

- **Frequency-related information:** We count the number of objects for each category, for example, rain or puddle. The existence or the number of objects can be linked to a few DAPR assessment scales such as *"No rain (Rain is present, No rain or other precipitation)"* or *"Puddle(s) (Number of puddles)"*.
- **Distance-related information:** To estimate the distance-related DAPR assessment scales such as *"standing in the puddle(s)" and "Lightning hit(s)"*, we calculate the distances among the identified objects, for example, between a person and a puddle or lightning and a person. The bounding box prediction from the detection model includes positional information, that is, 4-dimensional vector (x, y, w, h), where (x, y) is the bottom left coordinate and (w, h) is the width and height of the bounding box. Hence, the center coordinate of the bounding box prediction can be simply calculated as $((x + w)/2, (y + h)/2)$. By calculating the distance between the center coordinates of the bounding box of the objects, we estimate the positional relationship between the objects.
- **Area-related information:** To estimate the area-related DAPR scale such as *"Size of the figure (between 2 - 6 inches)"*, we compare the size of each object, where they are of the bounding box of an object can be easily calculated as $w * h$. Since the unit of the drawing canvas is a pixel, we convert the pixels-scale size to the inch-scale size.

3.3 An Explainable Dashboard

We developed a web-based dashboard to present the results and the evidence derived from Sections 3.1.2 and 3.2 to experts. We chose to present the results as a dashboard because it is known to be intuitive and easy to use [8]. As illustrated in Figures 2 and 3, to reduce the experts' cognitive load [62], the dashboard is designed by dividing tasks into (i) DAPR score-related information and (ii) supplementary information.

3.3.1 DAPR Score-Related Information.

- **Sketch recognition result:** As the detection model's output forms the foundation for sketch scoring, we offer a visual explanation of the object detection results to enhance explainability.
- **DAPR score table:** It shows the score of each DAPR scoring scale, including each coping resource and stress-related score, rather than simply showing the final DAPR score. The final DAPR score is calculated by subtracting the stress-related score from the coping resource score [6, 33]. By showing the score of each DAPR scoring scale, an expert can understand the derivation of the final DAPR score.
- **Stress-related and resource-related attributes:** The explanation about how the DAPR score scale is calculated based on the DAPR questionnaire is provided. The explanation includes the representative

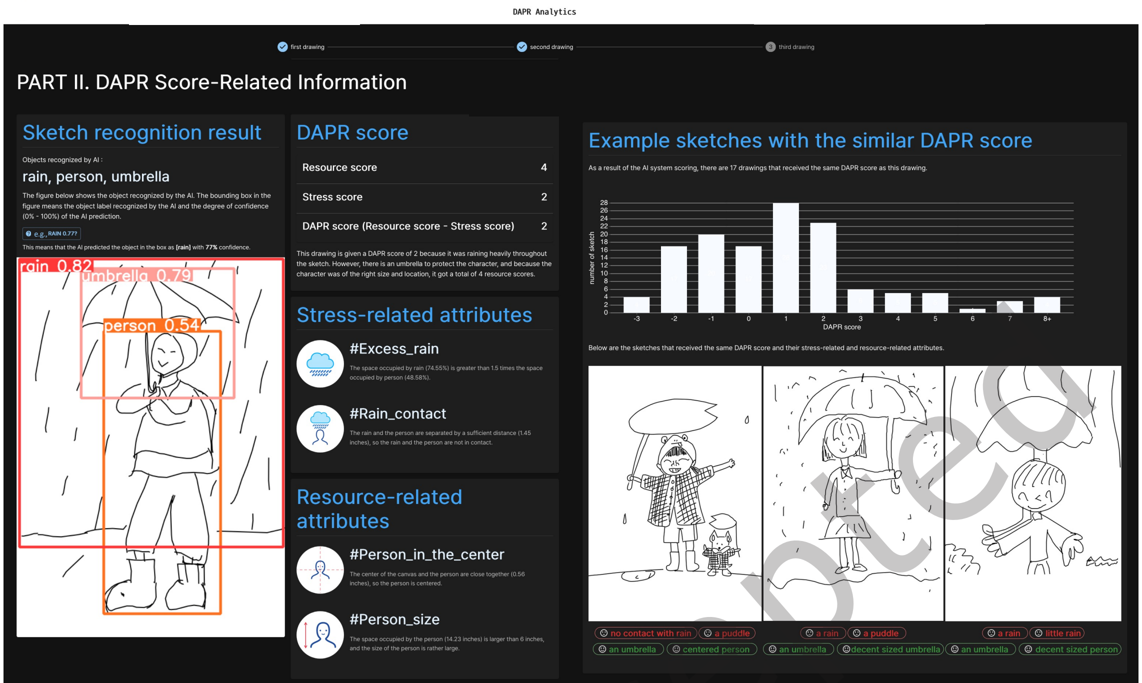

Fig. 2. User interface of DAPR Score-Related Information in an explainable dashboard, including sketch recognition result, DAPR score table, stress-related and resource-related attributes, and example sketches with similar DAPR score

keyword (e.g., '#Excess_rain' in the 'Stress-related attributes' board in Figure 2) of the result, and its text explanation, to provide an intuitive understanding of a given sketch.

- **Example sketches with a similar DAPR score:** Example sketches with the same DAPR score are provided with the following information: (1) the DAPR score distribution of the drawings in our dataset, collected in Section 3.1.1, and (2) the drawings with the same DAPR score with the keywords presenting their stress-related and resource-related attributes. This information provides references for the art therapists to refer to in the scoring process. In addition, it helps to intuitively interpret common psychological characteristics revealed in the sketches with the same DAPR score.

3.3.2 Supplementary Information.

- **Participant information:** The overall art therapy process and results can differ depending on the gender of the client [27]. Furthermore, the characteristics of the sketches, such as the length of stroke and the size of shapes, can differ from age [28]. Thus, we provide demographic information that can help to interpret the drawing results.
- **Drawing duration:** The information about the total drawing time of the given sketch is provided. Since psychiatric disorders such as schizophrenia affect the drawing time in the art therapy process [37], the drawing time information can help to understand the client's characteristics. This can reduce the workload of therapists and increase usability since measuring time through a time watch by an expert is not necessary anymore. A drawing duration can be easily calculated by subtracting the time of drawing the first stroke with the tablet pen from the time of submission of the drawing.

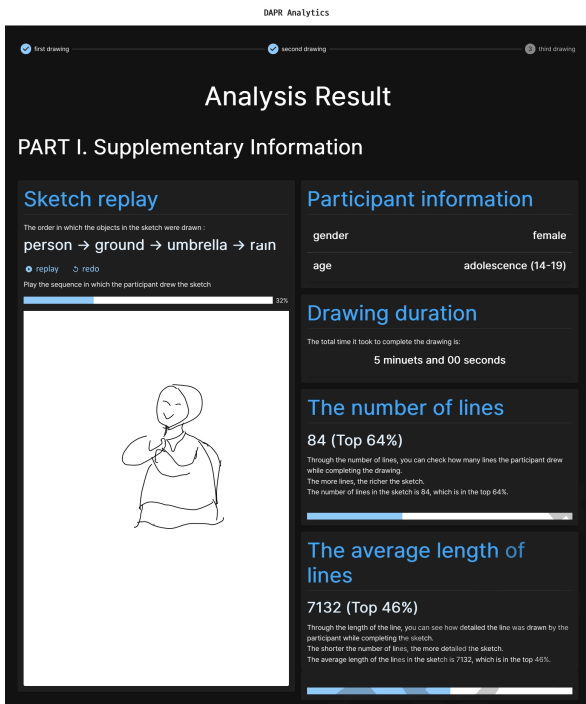

Fig. 3. User interface of Supplementary Information in an explainable dashboard, including participant information, drawing duration, sketch replay, and the number and the average length of lines.

- **Sketch replay:** In art therapy, therapists typically focus on the drawing process by a client [40]. Particularly, the drawing order that arranges the formatted introspection [42] is one of the important pieces of information. Therefore, a sketch replay function is provided to ensure that therapists can refer to the drawing process anytime, even if they cannot directly observe participants through this system. Thus far, the order in which the participants drew sketches on the tablet was saved, and when the play button was pressed, the drawing process was replayed as if it were a video. The therapist can review the drawing process at any time and as many times as needed.
- **The number and average length of lines:** In the DAPR assessment, line characteristics of a drawing such as line quality, pressure, shading, and stroke are important [9]. Hence, detailed stroke information while drawing the entire sketch is provided. The number of lines indicates the number of lines the participant drew while drawing the sketch, and the higher the number of lines, the richer the sketch was drawn. The average length of the lines shows how short and detailed the participants drew the lines. Each piece of information is given as a statistical chart that shows the ranking of a given sketch across all the sketches in our dataset to help the experts' understanding [1].

4 STUDY 1: USER EXPERIENCE ON ALPHADAPR

To examine user experience on *AlphaDAPR*, we conduct Study 1 using a user-centric framework, evaluating the intention to use an AI-based expert support system.

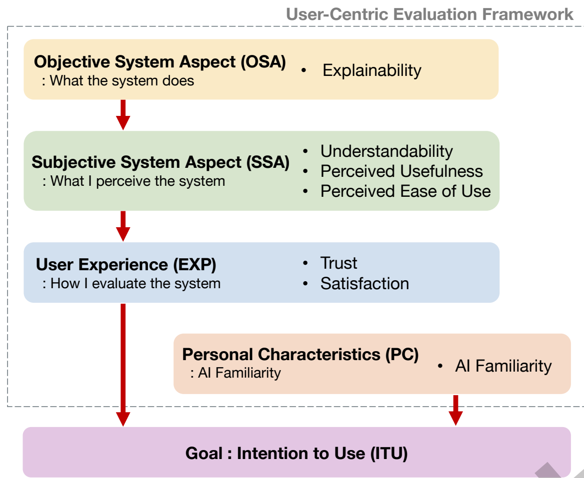

Fig. 4. The user-centric evaluation framework [31] with intention to use.

4.1 Research Method

4.1.1 User-Centric Evaluation Framework. User-centric evaluation frameworks are widely used to evaluate recommendation or decision support systems [31]. This framework connects the objective aspects (e.g., explainability) of the system with the subjective aspects (e.g., understandability and usefulness) of the system and the user experience (e.g., trust and satisfaction). The evaluation framework is similar to the technology acceptance model [16]; however, it considers a hedonic aspect of the user's experience in evaluating a system [31]. Hence, in Study 1, we adopted this framework to evaluate the intention to use the proposed AI-based expert support system for art therapists by considering user experience, which can be a mediator of intention to use.

Following the user-centric evaluation framework [31], we considered five factors to measure the intention to use the system: (i) objective system aspect (OSA), (ii) subjective system aspect (SSA), (iii) user experience (EXP), (iv) personal characteristics (PC), and (v) intention to use, which is the goal of the system. First, the *OSA* is the starting point for the framework. The explainability of a system (Explainability) is a measure of *OSA*. The *OSA* is then connected to the *SSA*. *SSA* represents how a user perceives the system, which measures whether the user understands well (Understandability), evaluates the system as useful (Perceived Usefulness), and accepts that easy to use (Perceived Ease of Use). *EXP* represents an evaluation of the system, such as trust and satisfaction, that is influenced by *SSA*. Additionally, we assume that the intention to use is affected by *PC*, which is measured by the extent to which a user is familiar with and trusts AI technology.

4.1.2 Experiment Procedure. Art therapy experts were recruited to conduct experiments to verify the effectiveness of the proposed system. Participants could connect to *AlphaDAPR* via personal laptops at desired locations (e.g., home and place of work) during the experiment. The participants initially signed their names and emails. An account was created based on this information and each user's data was stored in a user database. Hence, even if participants stopped using the system in the middle, they could continue to work based on the endpoint. Upon logging into their accounts, guidelines that included information on the experimental procedure were provided. Subsequently, the participants examined the results of the three sketches. Participants explored each analytical item (i.e., DAPR score-related and supplementary information) on the dashboard. After confirming that all three sketches had been examined, all participants completed the questionnaires. Of the 64 participants, 35 expressed

Demographic Variables	Frequency	Percent (%)
Sample size	64	100.0
Age		
21-30 years	13	20.3
31-40 years	27	42.2
41-50 years	19	29.7
51-60 years	5	7.8
Education		
Bachelor	10	15.6
Master	42	65.6
PhD	12	18.8
Art Therapist Work Experience		
No Experience	6	9.4
1-4 years	30	46.9
5-9 years	14	21.9
10-14 years	11	17.2
15-20 years	3	4.7

Table 2. Demographic profile of the participants.

their intention to participate in additional one-on-one interviews, of which eight participants were interviewed to obtain qualitative information on their experience of using the system. This study was approved by the Institute Review Board (Sungkyunkwan University[2] No. 2022-08-035).

4.1.3 Study Setup. In Study 1, we used three pre-selected sketches. We first sorted the collected 132 sketches by the number of strokes. We then selected one sketch from each of the 1st, 2nd and 3rd, and 4th quartiles in terms of the number of strokes, respectively. That is, three sketches were selected with a different number of strokes. Sixty-four art therapists participated in the system and answered 24 user experience questions, 6 personal characteristics questions, and 2 attention check questions [16]. The study was designed to be completed in 30 min, and each participant received $7 as monetary compensation.

- **Participants:** Experts with art therapy degrees were also recruited. A bachelor's degree was the minimum educational background required to qualify for participation in the experiment. Because *AlphaDAPR* was designed to run on a laptop, all participants participated in the experiment via their laptops.
  Similar to Guo et al. [16], two attention check questions like "Please select 'neutral' " were inserted among the 24 user experience questions to ensure the quality of the participant's responses. Among all participants, 58 had post-degree work experience in hospitals or centers, with an average work experience of 5.56 (M = 4.48) years. The average age of the participants was 38.75 (SD = 8.16). Table 2 summarizes the participants' detailed information.
- **Measures:** To examine the effectiveness and usability of *AlphaDAPR*, we measured participants' experiences through a quantitative evaluation using a questionnaire and a qualitative evaluation through interviews. All questions were rated on a 5-point agreement scale ranging from strongly disagree to

[2]https://www.skku.edu/eng/Research/industry/IRB.do

Considered aspects	Items	Factor loadings
Explainability [58]	• The textual information provided by this system helped me understand the decision-making process of the system.	0.669
	• The visualisation provided by this system helped me understand the decision-making process of the system.	0.764
Understandability [41]	• I understand how the system will assist me with decisions I have to make.	0.595
	• It is easy to follow what the system does.	0.581
	• I recognize what I should do to get the advice I need from the system the next time I use it.	0.915
Perceived Usefulness [11]	• The system enables me to accomplish tasks more quickly.	0.787
	• Using the system increases my diagnostic ability.	0.795
	• Using the system makes it easier to do my job.	0.774
	• Overall, I find the system useful in my job.	0.877
Perceived Ease of Use [11]	• It is easy for me to remember how to perform tasks using the system.	0.845
	• Overall, I find the system easy to use.	0.877
Trust [47]	• I believe the information that the system provides me.	0.937
	• This system is trustworthy.	0.896
Satisfaction [61]	• I am very satisfied with the information I receive from the system	0.841
	• All things considered, I am very satisfied with the system	0.944
AI Familiarity [59]	• I usually keep an eye on emerging technology products.	0.694
	• I always try out new technology products earlier compared to others.	0.919
	• I have had a deep understanding of artificial intelligence from long time ago.	0.646
Intention to Use [38] [46]	• Assuming that I have assess to the system, I intend to use it.	0.988
	• Assuming that I have assess to the system, I intend to use it frequently.	0.857

Table 3. Survey items with factor loadings.

strongly agree. All questionnaire items were translated from English to Korean and provided to participants.

As listed in Table 3, to examine the effects of explanation and explainability, we measured *explainability* using the two items in [58] (Cronbach's α = 0.746). *Understandability* was measured with three items reported in [41] (Cronbach's α = 0.731). *Perceived Usefulness* and *Perceived Ease of Use* were measured with two and four statements, respectively, developed by [11] (Cronbach's α = 0.880 and 0.850, respectively). *Trust* was measured with two items used in [47] (Cronbach's α = 0.916). *Satisfaction* was measured with three items from [61] (Cronbach's α = 0.895). *AI Familiarity* was measured with two parts; personal innovativeness in IT from [59] and understanding of AI (Cronbach's α = 0.787). Finally, to confirm that the participants were willing to continue using the system, their *Intention to Use* was measured using two statements derived from [38] and [46] (Cronbach's α = 0.896).

4.2 Quantitative Results

For quantitative evaluation, we used descriptive analysis, correlation analysis, and structural equation modeling (SEM). SPSS software version 25 and R version 4.2.1 were used for the analysis.

4.2.1 Descriptive Analysis. Table 4 summarizes the survey results for each variable. The participants reported that they were willing to use this system and satisfied with it. Over half of the participants (64.06%) reported

	M	SD	Percentage of positive responses
Intention to Use	3.80	0.91	64.06%
Explainability	3.97	0.76	73.44%
Understandability	4.09	0.74	79.69%
Perceived Usefulness	3.71	0.93	48.44%
Perceived Ease of Use	4.14	0.82	84.38%
Trust	3.43	0.87	40.63%
Satisfaction	3.52	0.83	46.88%
AI Familiarity	3.16	0.80	14.06%

Table 4. Mean, Standard Deviation, and Percentage of positive responses. Note that positive responses include 4 (agree) and 5 (strongly agree).

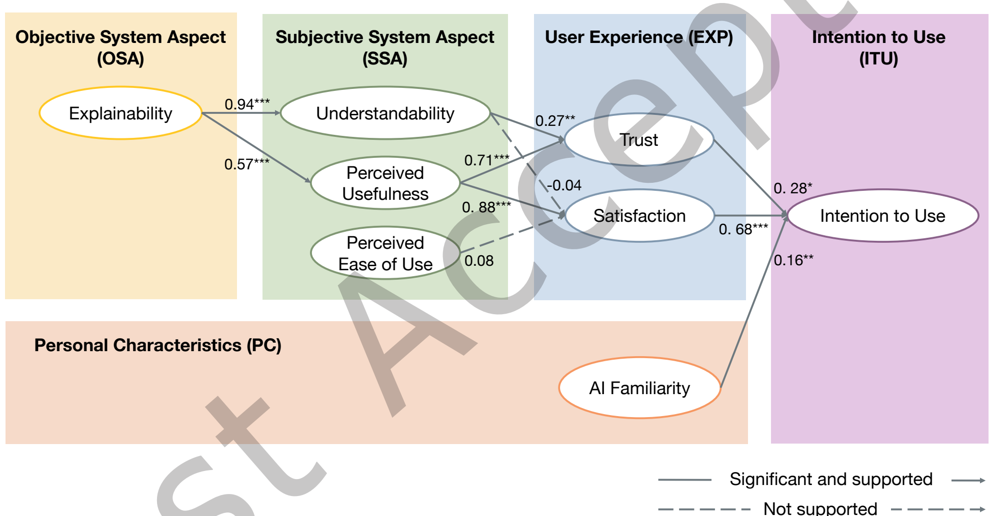

Fig. 5. The result of the structural equation modeling analysis. (Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.)

that they either intended or were eager to use it. In terms of explainability, 73.44% of the participants agreed that the explanation of the system was helpful for understanding (M = 3.97, SD = 0.76). Moreover, most of the participants reported that the system was understandable (M = 4.09, SD = 0.74), useful (M = 3.71, SD = 0.93), and easy to use (M = 4.14, SD = 0.82). Satisfaction and trust were assessed to investigate the user's experience and evaluate the system. In general, 46.88% of the participants answered that they were satisfied (M = 3.52, SD = 0.83), whereas trust in the system was slightly lower than in other items (M = 3.43, SD = 0.87). AI familiarity indicates how familiar a user is with the AI technology. Only 15.63% of the participants agreed (M = 3.34, SD = 0.74) that they were familiar with AI.

	1	2	3	4	5	6	7	8
1. Intention to Use								
2. Explainability	0.544^{***}							
3. Understandability	0.582^{***}	0.644^{***}						
4. Perceived Usefulness	0.817^{***}	0.379^{***}	0.455^{***}					
5. Perceived Ease of Use	0.552^{***}	0.576^{***}	0.693^{***}	0.500^{***}				
6. Satisfaction	0.788^{***}	0.450^{***}	0.435^{***}	0.799^{***}	0.420^{***}			
7. Trust	0.704^{***}	0.514^{***}	0.588^{***}	0.751^{***}	0.462^{***}	0.709^{***}		
8. AI Familiarity	-0.066	0.031	0.136	-0.160	0.091	-0.217	-0.194^{*}	

Table 5. Correlation between Intention to Use, Explainability, Understandability, Perceived Usefulness, Perceived Ease of Use, Satisfaction, Trust, and AI Familiarity. (Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.)

4.2.2 Analysis of Structural Equation Modeling (SEM). To investigate the determinants influencing the intention to utilize the system, we conducted an SEM analysis using R, grounded in the user-centric evaluation framework and the technology acceptance model. First, we create a correlation matrix, as listed in Table 5. Correlation analysis shows that the highest correlation is 0.817, which is less than 0.9; therefore, we confirm that there is no multicollinearity issue [43]. The model was validated using goodness of fit statistics. The measures include the Chi-square (p-value), root mean square error approximation (RMSEA), non-normed fit index (NNFI), and comparative fit index (CFI). The Chi-square statistic is significant, with $\chi^2 = 222.036$, p=0.002, df=157, χ^2/df=2.11 (<3). The RMSEA, whose value is less than 0.08 is considered a practical proof of a good fit [19] and was 0.080. In general, if the NNFI and CFI are greater than 0.9, then the model is considered suitable [21]. The NNFI, which is a modified version of the normed fit index (NFI) designed to reduce dependency on sample size [19], was 0.915. The comparative fit index (CFI was 0.930. Overall, these statistics indicate that the proposed model has an acceptable fit.

	System Functions	Frequency	% of Participants (n = 64)
DAPR Score-Related Information	Sketch recognition result	32	50.00
	DAPR score table	33	51.56
	Stress-related attributes	33	51.56
	Resource-related attributes	30	46.88
	Example sketches with the similar DAPR score	37	57.81
Supplementary Information	Participant information	44	68.75
	Sketch replay	55	85.94
	The number and average length of lines	41	64.06

Table 6. The functions of the *AlphaDAPR* that the participants recognized as useful (multiple choice).

As illustrated in Figure 5, most of the local paths were significant. *Explainability*, an explanation of the AI-related results provided by the system, was demonstrated to be a strong predictor for both *Understandability* ($\beta = 0.94$, $p < 0.001$) and *Perceived Usefulness* ($\beta = 0.57$, $p < 0.001$). In *SSA*, *Understandability* was significantly associated with *Trust* in the system ($\beta = 0.27$, $p < 0.01$), but did not have a significant impact on *Satisfaction*. *Perceived Usefulness* was shown as the very strong predictor for both *Trust* ($\beta = 0.71$, $p < 0.001$) and *Satisfaction* ($\beta = 0.88$, $p < 0.001$). However, the *Perceived Ease of Use* has no significant effect on *Satisfaction*. Both *Trust* and

Satisfaction positively impact users *Perceived Intention to Use* ($\beta = 0.28$, $p < 0.05$, and $\beta = 0.68$, $p < 0.001$). Finally, *AI familiarity* affected the *Intention to Use* the system ($\beta = 0.16$, $p < 0.01$). The results indicate that increasing trust and satisfaction in a system can increase the *Intention to Use*.

4.3 Qualitative Results

The results of the quantitative analysis demonstrate that most participants were willing to use the proposed system. In this section, we focus on how participants evaluated each part of the system and why. Thus far, we asked about the extent to which the participants perceived each part of the system, including the sketch replay, sketch recognition results, and DAPR score table.

Table 6 shows the functions of the system that the participants recognized as useful. As listed in Table 6, each piece of information (both DAPR score-related and supplementary) is regarded as useful, implying that *AlphaDAPR* has the potential to support art therapists. Supplementary information was created and provided by following the guidance of art therapists to include specific features that can help conduct a large-scale drawing assessment. Thus, not only DAPR score-related information but also Supplementary information received positive responses from art therapists.

For the qualitative investigation, we asked open questions at the end of the questionnaire as follows: *(i) which functions the participants liked and disliked when using the system, and (ii) suggestions to improve the system.* After the survey, we recruited volunteers from among the participants for an additional semi-structured interview with the above questions, and eight volunteers were interviewed. An additional interview was conducted for approximately 15-20 min, which was recorded and transcribed for analysis. To examine the participants' responses to the open questions and interviews, we organized the comments based on the related system parts and their intention to use the system. Subsequently, we categorized the responses into positive and negative attitudes. In this study, frequently observed patterns and themes were identified.

4.3.1 Benefits of an AI-based expert support system. Numerous participants answered that AI-provided information (e.g., the number and average length of lines and stress- and resource-related attributes) was useful. We first found that the participants reported that the proposed system could be effective in screening many drawings when an art therapist needs to analyze a large number of sketches, for example, in a school. One stated that *"Although the AI-provided DAPR score somewhat differed from the manually calculated DAPR score, I believe this system can be useful to reduce a therapist's time and burden when there are many drawings to review." (P21).*

Another benefit of the proposed system is that it provides objective and quantitative evidence for the assessment. P2 stated that *"I found the system helpful because it provides quantified assessment even for the ones that were used to be evaluated intuitively."* Furthermore, P10 answered that *"This system helps me come up with overall evaluation in a more objective fashion, particularly with the information about the line and the stress-related attribute... And I like the fact that I could get another opinion that I may be able to consult."* One (P48) indicated that the quantified analysis of drawings could enhance the trust of a client in the results of the drawing assessment, which may motivate participants to participate in art therapy.

As a newly introduced function for art therapists, example sketches with similar DAPR scores were recognized as useful by 57.81% of the participants when interpreting a sketch. P2 explained, *"This information provides criteria by presenting the sketches with similar results to interpret the client's sketch."* (P2). However, several participants expressed that 'example sketches with a similar DAPR score' might not be useful in deciding on the given sketch. P4 commented that *"The information on sketches with the same score could be helpful to enhance the trust of this system. However, because every client has a different background, the results of other clients' DAPR scores for interpreting the target drawing may not be comparable."* P23 commented that *"I could not agree that the presented sketches have the same score with the target client's sketch."*

4.3.2 *Intention to use the system.*

Numerous participants expressed their intention to use the system as assistance that can support understanding the client's sketch quantitatively. *"I trust the stress-related attributes, and this information is useful to support the analysis on the sketch. However, it could not provide a full understanding of the client's sketch because the system was analyzed only quantitatively."* (P25). When interpreting a client's sketch, an art therapist comprehensively considers multiple contexts, such as the visual characteristics of the sketch, the client's information, and how the client describes the sketch. P10 commented that *"This system mainly considers the visual aspect of the sketch-like an art therapist with relatively little experience. Hence, I suggest providing guidelines on how to use this system for interpreting drawings."*

Several participants made negative comments regarding the system. P9 commented on a reason not to use the system, *"For more accurately interpreting the result of the drawing assessment, the AI should recognize more fine-grained objects and detailed facial expressions."* In addition, P16 stated that *"I am suspicious about how AI can interpret a client's psychological state like an art therapist. I doubt the reliability of the system's analysis results."*

5 STUDY 2: THE EFFECTS OF A CLEAR COMMUNICATION ON THE ROLES OF BOTH AI AND USERS

Overall, the results in Study 1 indicated that art therapy experts tend to agree that *AlphaDAPR* can be a useful tool, especially with its objective and quantitative analysis functions. However, the level of trust was relatively low, and its impact on intention to use was not as substantial as we expected. This is rather surprising, given the fact that many previous studies showed that trust is an important factor in decision support systems [29, 45] and Kim et al. [29] even showed that trust is a more important factor than satisfaction in enhancing intention to use.

From the qualitative data from our in-depth interview in Study 1, we gathered evidence that the participants valued *AlphaDAPR*'s objective and quantitative analysis functions but questioned its qualitative and more in-depth analysis skills, which require comprehensive consideration of various factors such as background and contextual information. Moreover, several participants were concerned about overtrusting the AI-based system or even replacing human jobs despite its limitations. This helped us think that we may not have communicated clearly that our AI system was only a supportive tool that supplements human experts' final evaluation. We noticed that the collaborative roles between the machine and humans were implied in the current system but not explicitly communicated. The proposed system was useful for obtaining quantitative data (e.g., the number of strokes) quickly and efficiently, as recognized by most participants. We postulated that it is important to communicate more clearly that AI plays a supporting role by providing supplemental information (i.e., objective quantitative analysis), and human experts will make an overall evaluation with the help of AI so that users understand the distinct role of both humans and AI. In that way, users will not expect what AI cannot do (i.e., in-depth analysis with comprehensive evaluation) and will not get intimidated by the possibility of AI replacing their job, which possibly results in enhancing trust in AI.

For this reason, we designed Study 2 to examine if our postulation is correct by investigating whether providing explanations on clear definitions of AI and user roles enhances trust. Thus, we conducted an experiment to compare the following two groups: one with a clear explanation of the roles of AI and users and the other without such information.

5.1 Research Method

5.1.1 *Experiment Procedure.*

For an online experiment, twenty-six art therapy experts were recruited to investigate the effects of clear communication about the roles of AI and users on the users' perceived trust. Participants were randomly assigned to one of the two conditions: one with an explanation and one without. One group received a clear explanation of the roles of both AI and users based on an image-based explanation (see Figure 6) to make it easy to understand. The explanation highlighted that *AlphaDAPR* is a support system for art therapy experts. It also emphasized the roles of both AI and humans: *AlphaDAPR* supports quantitative analysis of

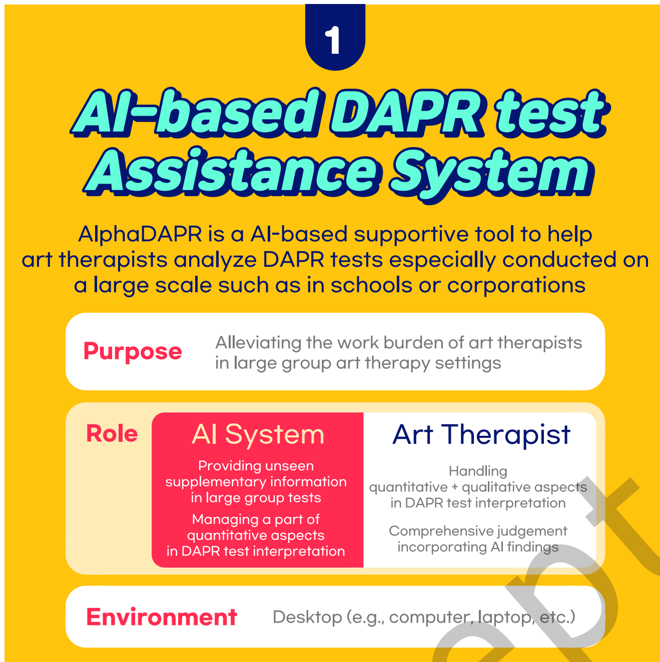

Fig. 6. An example of the used explanation that can clarify roles of *AlphaDAPR*

large-scale groups, while art therapists complete the final analysis considering the analysis results provided by the AI system as supplemental information. The other group, following the same experiment setup as Study 1 in Section 4.1.2, did not receive any clear explanation regarding their roles. Then, all the participants were asked to evaluate the three pre-selected sketches, examining each element of the DAPR score-related and supplementary information on the dashboard. Following a review of the three sketches, all participants completed the questionnaire.

5.1.2 Study Setup. Twenty-six art therapists participated in our system and answered 1 manipulation check question and 7 cognition-based trust questions. Each participant was paid $7 as compensation, and the experiment was structured to conclude within 30 min.

- **Participants:** We recruited experts with art therapy and bachelor's degrees as the minimum requisite educational background. Given that *AlphaDAPR* was designed to run on a laptop, all participants enlisted in the experiment utilized their laptops. The average age of participants was 39.0 (SD = 7.88). Among all 26 participants, 24 participants had work experience in art therapy, with an average work experience of 6.92 (SD = 4.21) years. Table 7 lists the information for each participant group. Group A, consisting of 13 participants, received a clear explanation of the roles of AI and its users, whereas Group B did not.
- **Measures:** To assess how the participants perceived *AlphaDAPR*, we incorporated a single manipulation check item: *"What do you think of the role of AI in the given system?"*. We then investigated the *trust* in the system, which was measured by a scale proposed by Madsen and Gregor [41]. Three cognition-based trust factors were measured: *perceived reliability*, *perceived technical competence*, and *perceived understandability*. *Perceived reliability* is in the usual sense of repeated and consistent functioning of the system, and *perceived technical competence* indicates that the system is perceived to perform the tasks accurately and correctly

Demographic Variables	Frequency		Percent (%)
Sample size	Group A 13	Group B 13	100.0
Age			
21-30 years	1	2	11.5
31-40 years	8	8	61.6
41-50 years	2	1	11.5
51-60 years	2	2	15.4
Education			
Bachelor	1	0	3.8
Master	9	10	73.1
PhD	3	3	23.1
Art Therapist Work Experience			
No Experience	0	2	7.7
1-4 years	5	4	34.6
5-9 years	6	3	34.6
10-14 years	2	3	19.3
15-20 years	0	1	3.8

Table 7. Demographic profile of the participants.

Factors	Items
Perceived Reliability	• This system performs reliably. • I can rely on the system to function properly. • The system analyzes problems consistently.
Perceived Technical Competence	• This system has sound knowledge about DAPR art therapy built into it. • The system correctly uses the provided information (client information).
Perceived Understandability	• I understand how this system will assist me with decisions I have to make in art therapy. • It is easy to follow how the system operates.

Table 8. Trust Items of the effectiveness of clarifying the roles.

	Topic (token percentage)	words
Group A	1 (60.4% of tokens)	assistant, numerical, system, therapist, possible, provide, art, tool, objective
	2 (39.6% of tokens)	sketches, objective, result, role
Group B	1 (53.3% of tokens)	assessment, evaluation, role, umbrella, size, resource, coping, location, element, degree
	2 (32.9% of tokens)	analysis, people, assessment, identification, basic, statistics
	3 (13.8% of tokens)	assistant, tool, role, assessment

Table 9. The identified topics in Group A and Group B.

based on the input information. *Perceived understandability* represents that the human supervisor or observer can form a mental model and predict future system behavior. Based on the selected factors, seven items were adopted to measure *trust*, as listed in Table 8. All items were translated from English to

	Group A		Group B		T-test	
	M	SD	M	SD	T	p
Perceived Reliability	3.87	0.44	3.74	0.31	0.86	0.40
Perceived Technical Competence	3.85	0.38	3.46	0.32	2.81	0.01
Perceived Understandability	4.31	0.48	3.92	0.31	2.50	0.02
Average	4.05	0.34	3.71	0.17	2.46	0.02

Table 10. Mean, Standard Deviation, and Percentage of positive responses. Note that positive responses include 4 (agree) and 5 (strongly agree).

Korean for the participants, and a 5-point agreement scale ranging from strongly disagree to strongly agree was used.

5.2 Results

5.2.1 Manipulation Check. The results of the manipulation check revealed how participants recognized the role of *AlphaDAPR*. One group that received a clear explanation of the roles of AI and users (Group A) reported the role of AI as 'an assistant tool,' 'a support system,' and 'a tool for quantitative analysis.' For instance, one participant replied that *AlphaDAPR* plays *"an assistant role by providing results to help art therapists comprehensively analyze and evaluate."* However, the other group, without a clear explanation about the roles (Group B), recognized the role of AI as 'a role in evaluating the quality of pictures,' 'a basic component identification tool,' and 'a tool for quick information acquisition and results organization.' One participant commented on its role as *"the role of recording the drawing process and evaluating its quality."*

For an in-depth analysis, we utilized an LDA-based topic modeling approach [23] for each of the two groups. We identified two topics in Group A and three topics in Group B based on the highest topic coherence, as summarized in Table 9. Note that all the identified topics in Table 9 are manually checked to ensure accuracy and eliminate noise. As a result, in Group A, topic 1 involves terms related to supporting systems, and topic 2 is characterized by keywords related to drawing assessment. On the other hand, topic 3 in Group B has fewer tokens associated with supporting systems (e.g., assistant, tool). Also, topics 1 and 2 in Group B contain keywords related to DAPR attributes.

Overall, the results show that Group A (with an explicit explanation regarding the roles of AI and users) comprehended the role of AI as "a supportive tool for the quantitative analysis of art therapists," as directed before the experiment. On the other hand, Group B (without an explicit explanation) recognized that the proposed system was "a tool for result organization."

5.2.2 Trust. Table 10 shows the result of the comparison between the group that received a clear explanation about the roles of AI and users (Group A) and the group that did not (Group B). Group A showed a higher trust (M = 4.05, SD = 0.34) compared to Group B (M = 3.71, SD = 0.17). A Student's T-test revealed that there is a significant difference between the group with a clear explanation and one without it on trust (t = 2.46, $p < 0.05$, df = 24.0). This suggests that transparent delineation of the roles of AI and users enhances participants' trust in comparison to scenarios devoid of any explanation.

We found that the group with clear explanations scored higher on all the sub-dimensions of trust as well including *reliability* (t = 0.86, $p < 0.5$, df = 24.0), *technical competence* (t = 2.81, $p < 0.05$, df = 24.0), and *understandability* (t = 2.50, $p < 0.05$, df = 24.0). This indicates that clear communication enhances trust in the technical proficiency and accuracy of a system.

6 DISCUSSION

In this section, we discuss the improvement of the intention to use the system. We also describe the practical implications and limitations of the *AlphaDAPR*.

6.1 Trust, Intention to Use, and Clear Communication about the Roles of AI and Users

In this study, we introduce an AI-based system to support art therapists and psychologists by interpreting drawing assessment results. To develop this system, we applied the concept of explainable AI. Our findings in Study 1 highlight the importance of AI-based expert support systems and decision-making. The SEM analysis reveals that the paths from explainability, usefulness, trust, satisfaction, and intention to use are significant, implying that explainability is crucial in increasing the intention to use the proposed system.

However, satisfaction was a significant factor that directly affected the intention to use. This result aligns with [20], which showed that satisfaction is essential for raising the intention to use. On the other hand, trust partially or limitedly affects the intention to use, unlike many other studies [29, 45] showed that trust is an important factor in designing a decision support system. We conjectured that the lack of clear communication was the underlying cause. To validate this, in Study 2, we conducted an experiment to compare two groups who were given a clear explanation about the roles of AI and users or not. The results demonstrated that when we precisely defined the roles of AI and users, there was a notable increase in trust among art therapists regarding the system. Furthermore, the manipulation check results confirmed that misunderstandings regarding the system's purpose can arise when the roles of the AI and users are not explicitly provided. This implies that the absence of clear explanations about the system allows participants to freely interpret the roles of AI and users, potentially influencing the unexpected outcomes.

This serves as corroborative evidence for prior research findings [35, 44] and underscores the significance of accurately delineating the roles of AI in systems intended to support users. It is imperative to unequivocally convey that the final decision rests with the human user, supported by the AI system. This clarification ensures that users neither feel unduly threatened nor harbor misconceptions about AI superseding their roles. Clear communication about roles will also help users evaluate the AI system based only on what it is supposed to do (e.g., quantitative analysis) and not on what humans are supposed to do (e.g., qualitative in-depth analysis).

The importance of clear communication about roles, particularly in human-AI collaboration, has not been stressed sufficiently and largely unveiled in the previous literature, while emphasis has been placed on explainable AI. While we employed a straightforward approach, specifically using a simple image-based explanation to clarify the roles of AI and users, we believe that future studies should further investigate the importance of clear communication regarding the roles of AI and humans to provide more helpful guidelines for designing AI-based systems.

6.2 Practical Implications

When an art therapist decides on a client's drawing in the drawing assessment, he/she utilizes multiple resources, from quantitative to qualitative information. To follow the art therapy practice, we designed a dashboard for both quantitative and qualitative information on the drawing and client on a single page. As listed in Table 6, 68.53% and 85.94% of the participants liked the 'Participant information' and the 'Sketch replay' function in our dashboard, respectively. These results imply that providing sufficient information along with AI-provided information to support the decision is useful, particularly when an art therapist cannot fully participate in the entire drawing assessment process because of a non-face-to-face situation or a large-group session. For example, one art therapist in our interview suggested that presenting the results of post-hoc questions about the drawing is also helpful.

Alufaisan et al. [4] showed that the accuracy of the AI model is an important factor in using the AI-based system. Obtaining a sufficient number of training sketches is essential for improving the accuracy of the sketch-recognition model. However, collecting large-scale data for assessments is challenging. To solve this problem, we developed an automated drawing analysis model using an existing sketch dataset and synthesized the data. The performance of Yolo-v5 [12] in terms of AP increased when both the collected and synthesized sketches were used for training. These results show that utilizing methods for synthesizing data along with collected sketches is useful for improving the accuracy of the sketch recognition model. Although state-of-the-art sketch recognition models achieve considerable accuracy, their predictive results can be inaccurate or overlook fine-grained points, unlike human art therapists. Therefore, providing a human-in-loop feedback process can be useful, where an art therapist manually revises disagreeable predictive results from the AI and obtains updated information related to the predictive results, that is, DAPR score and stress- and resource-related attributes.

6.3 Limitations & Future Directions

This study has several limitations. First, participants were recruited using snowball sampling from a particular therapist; therefore, this study may not be representative of general art therapists. Therefore, further work requires a group with a more diverse environment (e.g., race). Second, the small sample size was a limitation. Although the pathways in the SEM analysis were significant, some pathways may have been insignificant because of relatively small sample sizes. Thus, our future work will include the recruitment of numerous art therapists to validate the paths identified in our study. Third, we collected digital drawings using a computer tablet to build the DAPR dataset; pen pressure was not considered in this process. However, pen pressure is also related to line quality and can provide useful information. Hence, in future work, we plan to analyze drawings that reflect pen pressure. Finally, this study aims to introduce a new AI-based expert support system for conducting large-scale drawing assessments and collecting opinions from art therapists. Our future work will include (i) optimizing the proposed system based on the lessons learned from our study, and (ii) validating the effectiveness of *AlphaDAPR* based on a controlled experiment.

7 CONCLUSION

To the best of our knowledge, this is the first attempt to propose an automatic analysis system for DAPR assessment, *AlphaDAPR*, which can support art therapists. *AlphaDAPR* can save art therapists time and money by analyzing clients' drawings and offering interpretable results. To increase the intention to use the proposed system, *AlphaDAPR* is developed with explainability. As a result, 64.06% of the participants indicated a willingness to use the proposed system in Study 1. Also, the results of structural equation modeling demonstrated that explainability affected perceived usefulness, trust, satisfaction, and intention to use. While we confirmed positive outcomes regarding *AlphaDAPR* in Study 1, trust was measured relatively low. We postulated that this is due to a lack of clear communication about the roles of AI and users. Therefore, in Study 2, we compared the trust of two groups who are with and without the clear explanation on the roles of AI and users. The results emphasized the importance of clear communication in users' trust. We believe that *AlphaDAPR* with clear explanation about the roles can be used as an effective assistant tool to reduce the heavy workload of art therapists, particularly in large-scale groups such as schools or companies.

ACKNOWLEDGMENTS

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Graduate School of Metaverse Convergence support program(IITP-2024-RS-2023-00254129) and the ITRC(Information Technology Research Center) support program(RS-2024-00436936) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–18.

[2] Lewis Aiken. 1999. Personality assessment: Methods & practices (3rd ed. rev.). (1999).

[3] Brian Allen and Chriscelyn Tussey. 2012. Can projective drawings detect if a child experienced sexual or physical abuse? A systematic review of the controlled research. *Trauma, violence, & abuse* 13, 2 (2012), 97–111.

[4] Yasmeen Alufaisan, Laura R Marusich, Jonathan Z Bakdash, Yan Zhou, and Murat Kantarcioglu. 2021. Does explainable artificial intelligence improve human decision-making?. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 6618–6626.

[5] Galit Amir and Rachel Lev-Wiesel. 2007. Dissociation as depicted in the traumatic event drawings of child sexual abuse survivors: A preliminary study. *The arts in psychotherapy* 34, 2 (2007), 114–123.

[6] Pui Kwan Au. 2020. *An application of FEATS scoring system in Draw-A Person-in-the-Rain (DAPR): Distinguishing depression, anxiety, and stress by projective drawing*. Ph. D. Dissertation. Hong Kong: Hong Kong Shue Yan University.

[7] Matthias Baldauf, Peter Fröehlich, and Rainer Endl. 2020. Trust me, I'ma doctor–user perceptions of AI-driven apps for mobile health diagnosis. In *19th International Conference on Mobile and Ubiquitous Multimedia*. 167–178.

[8] Enrico Bunde. 2021. AI-Assisted and explainable hate speech detection for social media moderators–A design science approach. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. 1264.

[9] Suellen M Carney. 1992. *Draw a person in the rain: A comparison of levels of stress and depression among adolescents*. Pace University.

[10] Shuqing Chen, Daniel Stromer, Harb Alnasser Alabdalrahim, Stefan Schwab, Markus Weih, and Andreas Maier. 2020. Automatic dementia screening and scoring by applying deep learning on clock-drawing tests. *Scientific Reports* 10, 1 (2020), 1–11.

[11] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.

[12] Glenn Jocher et. al. 2022. ultralytics/yolov5: v6.1 - YOLOv5n 'Small' models, Roboflow integration, TensorFlow export, OpenCV DNN support. https://doi.org/10.5281/zenodo.5563715

[13] Mr C Fairhurst, T Linnell, Stephanie Glenat, RM Guest, Laurent Heutte, and Thierry Paquet. 2008. Developing a generic approach to online automated analysis of writing and drawing tests in clinical patient profiling. *Behavior Research Methods* 40, 1 (2008), 290–303.

[14] Adam Graves, Leslie Jones, and Frances F Kaplan. 2013. Draw-a-Person-in-the-Rain: Does geographic location matter? *Art Therapy* 30, 3 (2013), 107–113.

[15] Simeng Gu, Yige Liu, Fei Liang, Rou Feng, Yawen Li, Guorui Liu, Mengdan Gao, Wei Liu, Fushun Wang, and Jason H Huang. 2020. Screening depressive disorders with tree-drawing test. *Frontiers in psychology* 11 (2020), 1446.

[16] Lijie Guo, Elizabeth M Daly, Oznur Alkan, Massimiliano Mattetti, Owen Cornec, and Bart Knijnenburg. 2022. Building Trust in Interactive Machine Learning via User Contributed Interpretable Rules. In *27th International Conference on Intelligent User Interfaces*. 537–548.

[17] Emanuel F Hammer. 1958. *The clinical application of projective drawings.* Charles C Thomas.

[18] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 2961–2969.

[19] Daire Hooper, Joseph Coughlan, and Michael R Mullen. 2008. Structural equation modelling: Guidelines for determining model fit. *Electronic journal of business research methods* 6, 1 (2008), pp53–60.

[20] Kuo-Lun Hsiao and Chia-Chen Chen. 2021. What drives continuance intention to use a food-ordering chatbot? An examination of trust and satisfaction. *Library Hi Tech* (2021).

[21] Yu-Kai Huang. 2010. The effect of airline service quality on passengers' behavioural intentions using SERVQUAL scores: A Taiwan case study. *Journal of the Eastern Asia Society for Transportation Studies* 8 (2010), 2330–2343.

[22] Chiara Ionio and Eleonora Mascheroni. 2021. Psychological well-being and graphic representations of self in child victims of violence. *The Arts in Psychotherapy* 72 (2021), 101740.

[23] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications* 78 (2019), 15169–15211.

[24] Hongchao Jiang, Yanci Zhang, Zhiwei Zeng, Jun Ji, Yu Wang, Ying Chi, and Chunyan Miao. 2021. Mobile-based Clock Drawing Test for Detecting Early Signs of Dementia. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 16048–16050.

[25] Juliet Jue and Jung Hee Ha. 2019. The Person-in-the-Rain Drawing Test as an Assessment of Soldiers' Army Life Adjustment and Resilience. *Psychology* 10, 8 (2019).

[26] Jiwon Kang, Jiwon Kim, Migyeong Yang, Chaehee Park, Taeeun Kim, Hayeon Song, and Jinyoung Han. 2024. SceneDAPR: A Scene-Level Free-Hand Drawing Dataset for Web-based Psychological Drawing Assessment. In *Proceedings of the ACM Web Conference 2024*.

[27] Daiki Kato and Miyako Morita. 2009. Form, content, and gender differences in Lego® block creations by Japanese adolescents. *Art Therapy* 26, 4 (2009), 181–186.

[28] Hong-hoe Kim, Paul Taele, Stephanie Valentine, Erin McTigue, and Tracy Hammond. 2013. KimCHI: a sketch-based developmental skill classifier to enhance pen-driven educational interfaces for children. In *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling*. 33–42.

[29] Jeoungkun Kim, Soongeun Hong, Jinyoung Min, and Heeseok Lee. 2011. Antecedents of application service continuance: A synthesis of satisfaction and trust. *Expert Systems with applications* 38, 8 (2011), 9530–9542.

[30] Limor Kissos, Limor Goldner, Moshe Butman, Niv Eliyahu, and Rachel Lev-Wiesel. 2020. Can artificial intelligence achieve human-level performance? A pilot study of childhood sexual abuse detection in self-figure drawings. *Child Abuse & Neglect* 109 (2020), 104755.

[31] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22, 4 (2012), 441–504.

[32] Kate Kravits, Randi McAllister-Black, Marcia Grant, and Christina Kirk. 2010. Self-care strategies for nurses: A psycho-educational intervention for stress reduction and the prevention of burnout. *Applied Nursing Research* 23, 3 (2010), 130–138.

[33] Christine P Krom. 2000. *Hospice nurses and the palliative care environment: Indicators of stress and coping in the Draw-a-Person-in-the-Rain test.* Ph. D. Dissertation. Albertus Magnus College.

[34] Heidi S Lack. 1997. *The person-in-the-rain projective drawing as a measure of children's coping capacity: a concurrent validity study using rorschach, psychiatric, and life history variables.* California School of Professional Psychology-Berkeley/Alameda. 156–174 pages.

[35] Mengqi Liao and S Shyam Sundar. 2021. How should AI systems talk to users when collecting their personal information? Effects of role framing and self-referencing on human-AI interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[36] Eva Fishell Lichtenberg. 2014. Draw-A-Person-in-the Rain Test. In *Drawings in assessment and psychotherapy: Research and application*, Leonard Handler and A. D. Thomas (Eds.). Routledge/Taylor & Francis Group., 164–183.

[37] Yu Shiou Lin, Peter Hartwich, Annemarie Wolff, Mehrshad Golesorkhi, and Georg Northoff. 2020. The self in art therapy–brain-based assessment of the drawing process. *Medical Hypotheses* 138 (2020), 109596.

[38] Pin Luarn and Hsin-Hui Lin. 2005. Toward an understanding of the behavioral intention to use mobile banking. *Computers in human behavior* 21, 6 (2005), 873–891.

[39] Karen Machover. 1949. Personality projection in the drawing of the human figure: A method of personality investigation. (1949).

[40] Natalie Mackenzie. 2013. A brief exploration of the role of dramatherapy within a multi-modal arts therapy approach to working with children aged 4–14 years impacted by trauma. *Dramatherapy* 35, 2 (2013), 131–139.

[41] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, Vol. 53. Citeseer, 6–8.

[42] A Meneghetti. 2004. Ontopsychology handbook. *Ontopsicologia Editrice, Roma* (2004).

[43] Habshah Midi, Saroje Kumar Sarkar, and Sohel Rana. 2010. Collinearity diagnostics of binary logistic regression model. *Journal of interdisciplinary mathematics* 13, 3 (2010), 253–267.

[44] Tanya Nazaretsky, Mutlu Cukurova, and Giora Alexandron. 2022. An instrument for measuring teachers' trust in AI-based educational technology. In *LAK22: 12th international learning analytics and knowledge conference*. 56–66.

[45] Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *CHI Conference on Human Factors in Computing Systems*. 1–9.

[46] Sung Youl Park. 2009. An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. *Journal of Educational Technology & Society* 12, 3 (2009), 150–162.

[47] Robin Pennington, H Dixon Wilcox, and Varun Grover. 2003. The role of system trust in business-to-consumer transactions. *Journal of management information systems* 20, 3 (2003), 197–226.

[48] Ochilbek Rakhmanov, Nwojo Nnanna Agwu, and Steve Adeshina. 2020. Experimentation on hand drawn sketches by children to classify Draw-a-Person test images in psychology. In *The Thirty-Third International Flairs Conference*.

[49] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28 (2015).

[50] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[51] Woosuk Seo, Joonyoung Jun, Minki Chun, Hyeonhak Jeong, Sungmin Na, Woohyun Cho, Saeri Kim, and Hyunggu Jung. 2022. Toward an AI-assisted Assessment Tool to Support Online Art Therapy Practices: A Pilot Study. In *Proceedings of 20th European Conference on Computer-Supported Cooperative Work*. European Society for Socially Embedded Technologies (EUSSET).

[52] Rania Shibl, Meredith Lawley, and Justin Debuse. 2013. Factors influencing decision support system acceptance. *Decision Support Systems* 54, 2 (2013), 953–961.

[53] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551.

[54] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[55] Donghua Tao. 2009. Intention to use and actual use of electronic information resources: further exploring Technology Acceptance Model (TAM). In *AMIA Annual Symposium Proceedings*, Vol. 2009. American Medical Informatics Association, 629.
[56] JS Verinis, EF Lichtenberg, and L Henrich. 1974. The Draw-a-Person in the rain technique: Its relationship to diagnostic category and other personality indicators. *Journal of Clinical Psychology* (1974).
[57] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.
[58] Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, and Elisabeth André. 2021. "Let me explain!": exploring the potential of virtual agents in explainable AI interaction design. *Journal on Multimodal User Interfaces* 15, 2 (2021), 87–98.
[59] Fan Wenjuan, Jingnan Liu, Zhu Shuwan, and Panos M Pardalos. 2020. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research* 294, 1-2 (2020), 567–592.
[60] Lisa R Willis, Stephen P Joy, and Donna H Kaiser. 2010. Draw-a-Person-in-the-Rain as an assessment of stress and coping resources. *The Arts in Psychotherapy* 37, 3 (2010), 233–239.
[61] Barbara H Wixom and Peter A Todd. 2005. A theoretical integration of user satisfaction and technology acceptance. *Information systems research* 16, 1 (2005), 85–102.
[62] Minfan Zhang, Daniel Ehrmann, Mjaye Mazwi, Danny Eytan, Marzyeh Ghassemi, and Fanny Chevalier. 2022. Get To The Point! Problem-Based Curated Data Views To Augment Care For Critically Ill Patients. In *CHI Conference on Human Factors in Computing Systems*. 1–13.