



Article

# Emotion Recognition in Human–Robot Interaction Using the NAO Robot

Iro Athina Valagkouti <sup>1</sup>, Christos Troussas <sup>1,\*</sup> , Akrivi Krouska <sup>1</sup> , Michalis Feidakis <sup>2</sup> and Cleo Sgouropoulou <sup>1</sup>

<sup>1</sup> Department of Informatics and Computer Engineering, University of West Attica, 12243 Egaleo, Greece; cs161082@uniwa.gr (I.A.V.); akrouska@uniwa.gr (A.K.); csgouro@uniwa.gr (C.S.)

<sup>2</sup> Department of Electrical and Electronics Engineering, University of West Attica, 12243 Egaleo, Greece; m.feidakis@uniwa.gr

\* Correspondence: ctrouss@uniwa.gr

**Abstract:** Affective computing can be implemented across many fields in order to provide a unique experience by tailoring services and products according to each person’s needs and interests. More specifically, digital learning and robotics in education can benefit from affective computing with a redesign of the curriculum’s contents based on students’ emotions during teaching. This key feature is observed during traditional learning methods, and robot tutors are adapting to it gradually. Following this trend, this work focused on creating a game that aims to raise environmental awareness by using the social robot NAO as a conversation agent. This quiz-like game supports emotion recognition with DeepFace, allowing users to review their answers if a negative emotion is detected. A version of this game was tested during real-life circumstances and produced favorable results, both for emotion analysis and overall user enjoyment.

**Keywords:** emotion analysis; robotics in education; NAO robot; tailored feedback



**Citation:** Valagkouti, I.A.; Troussas, C.; Krouska, A.; Feidakis, M.; Sgouropoulou, C. Emotion Recognition in Human–Robot Interaction Using the NAO Robot. *Computers* **2022**, *11*, 72. <https://doi.org/10.3390/computers11050072>

Academic Editors: Ioana Andreea Stefan and Samir Garbaya

Received: 13 April 2022

Accepted: 30 April 2022

Published: 2 May 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Affective computing as defined by Picard (1997) is “computing that relates to, arises from, or deliberately influences emotions”. More specifically, this research field focuses on implementing the recognition, interpretation, and simulation of human emotions in machines [1–5]. Using sensors, cameras, and microphones, this technology can identify the emotional state of the user and adapt to it by performing a set of predetermined actions that fit the current situation.

Owing to affective computing, many fields, including education, can provide improved services by interpreting human emotion [6,7]. The ability to detect a person’s current emotional state and respond to their feelings is very important in services such as e-learning applications, where feelings, emotions, or mood can determine the delivery of the suggested content or change product features. With the use of robotics in education—a learning technique that enhances the learning process via the use of robots—affective computing can tailor the entire learning experience to the specific needs of each student. According to studies, such learning techniques produce better results than traditional face-to-face teaching across all educational stages [8–11].

Robotics provides innovative solutions and a wide range of applications that can be successfully used in educational settings. The interactive and tactile nature of robotics renders it capable of playing a leading role and offering great opportunities for teamwork and collaboration. Through robotics, learners can explore multidisciplinary fields and learn any course in an all-in-one learning experience. This experience involves great academic results in a pleasant environment where exogenous factors affecting learning or emotions can be taken into account.

Robotics can offer a fertile ground to learners to improve their skills and interests concerning technological issues. Interacting with robots in education stimulates learners

to strengthen their technical intuition and highlights purposeful, problem-based learning through the combination and utilization of their knowledge. Robotics in education ameliorates learners' fundamental skills, such as higher order thinking, problem solving, logical and analytical reasoning, as well as computational thinking.

As mentioned above, robotic systems can offer an optimal learning experience as they can provide autonomy and flexibility under a variety of circumstances or even simulate empathy. One such example is the humanoid NAO robot [12], which can perceive the emotional state of people as well as the environment's ambiance [13]. The robot utilizes various extractors and memory keys (e.g., head angles and smile information) to determine emotional reactions and classifies them as "positive", "neutral", "negative", and "unknown". However, it offers neither sufficient customization (e.g., the addition of more emotions) nor the option to use a custom face detection/recognition solution [12]. There are also limitations that hinder its performance, such as intense light sources, the use of glasses, and too-subtle facial expressions. Moreover, the NAO model does not yet support speech emotion recognition [13].

The objective of this paper was to research a lightweight and easy setup solution for a third-party machine learning model to interact with the NAO robot or its simulator in order to determine user emotion adequately. This was conducted by using a small-scale game-like quiz. The choices in this game provide different outcomes, but given the emotional state of the user, it can allow for re-evaluation of a given answer. In this work, we utilized NAO's cameras to capture images and a computer, acting as a server by hosting DeepFace [14], a face attribute analyzer. We used Python sockets to establish communication with the NAO robot in order to transfer data between the robot and the server and the Python Pickle module [15] to allow object serialization and reconstruction. The evaluation of the project was conducted under real-life circumstances, offering favorable results. It needs to be noted that the project can also be deployed on Choregraphe's simulator [16].

The novelty of this research lies in the incorporation of robotics in education and, specifically, the employment of the NAO robot. Moreover, this design is programmed to recognize users' emotions, and by offering a well-structured quiz-like game, it aims to promote users' environmental awareness. All these techniques smoothly function to provide an optimal user experience with the greatest pedagogical results. To the best of our knowledge, our presented approach has not been employed in the literature so far.

This paper is organized as follows. In Section 2, an overview of related applications and solutions involving social robots is presented. In Section 3, the overall functionality is described along with DeepFace and the robot's behavior. Section 4 consists of the proposed architecture, presenting the socket usage and NAO's conversation tool, QiChat [17]. Section 5 is dedicated to our results and evaluations. Section 6 includes the conclusion, the observed limitations, and further improvements.

## 2. Related Work

There is a growing interest in the use of social robots in the field of human-robot interaction (HRI). These robots range from being educational tools to means for treating patients, especially when used as assistant robots. A very promising candidate for HRI is the NAO robot, which is featured in many studies, especially when social interaction plays a crucial role [18–21]. In [18], the authors focused on autism spectrum disorder patients (ASD) and how their quality of life can be improved with the use of a social robot at an early age. In their study, they incorporated a neural network model that was able to generate meaningful answers that differed from the Seq2Seq model and the NAO social robot. Their experiments showed that the model surpassed the Seq2Seq model as well as the GAN-based one in both automatic and human evaluation. In [19], the authors proposed a face detection method in order to track ASD children's faces during robot-assisted therapy sessions. The authors noted that social robots were generating positive reactions from children with ASD, and they aimed to use NAO as an agent in order to measure the concentration level of children during a 30 min interaction session, which was supervised

by human doctors. In [21], a one-month experiment was presented, which was carried out in a home environment, featuring a social robot and a caregiver; the robot engaged in activities with children diagnosed with ASD. The activities featured emotional storytelling, perspective taking, and sequencing in order to increase the social communication skills of children. The results indicated that after the experiment, the children improved their joint attention skills and communicated more without needing the presence of the robot. In [20], the NAO robot was used as an assistant for elderly people. It was tested under a smart home robot system by eight older people in five real-world long-term and short-term scenarios. The humanoid robot was able to gain the participant's trust, and during the long-term evaluation, the results showed that the participants developed an emotional connection with the robot, although the overall entertainment gained from the interaction with the robot diminished.

A lot of studies featuring NAO combined it with emotion recognition [22–26]. Each approach had different objectives that were accomplished in different ways. The method proposed in [23] required the programming of a new, lighter version of the NAOqi library [27] for compatibility reasons and aimed at recognizing facial expressions in children. The one presented in [24] utilized the PEPPER robot and identified patients' expressions in order to determine their mental state by transferring the captured images to a local computer using a Python implementation of SSH. In [25], a four-stage emotion analysis algorithm was presented along with its application in NAO, focusing on detecting happiness, sadness, or anger, producing favorable results. The work presented in [22] aimed at recognizing human emotion with the NAO robot by using a pretrained convolutional neural network to classify expressions and later compare them with human experts' recognitions. The results of this experiment suggested that NAO is an adequate candidate for human–robot interaction, both for recognizing emotion and entertaining people. Finally, in [26], a facial expression recognition model was proposed that enhances NAO's facial expression perception; the model was integrated into the NAO SDK. This allowed for a fast facial expression classification implementation with very high accuracy scores for happy and sad expressions.

In the field of robotics in learning, robot tutees have been used in various research investigations focusing on different learning goals and demonstrating positive effects. In [28], the authors' main goal was to teach waste recycling to children with the use of a social robot. This was achieved by designing a serious game with the social robot PEPPER [29], which was able to detect and classify the waste material with a convolutional neural network. Other learning goals included teaching numeracy in elementary schools [30] and learning sign language [31] with the assistance of a robot via game-inspired learning techniques. Social robots can also be used for more abstract concepts, such as providing aid to elementary school children in order to retain their creativity [32] or teaching emotion recognition to children with autism [33]. In [32], the authors involved a verbal creativity task to generate titles for abstract images (Doodle Creativity Game) and showed that it encouraged children to think creatively. In [33], the authors used the NAO robot in a guessing game where children were tasked with deducting which emotion NAO was physically imitating. All of the abovementioned studies showed interesting results in learning effectiveness and the improvement of cognitive skills, and they prolonged the engagement of the children in the learning process. Finally, in [34], a fully autonomous social robot platform, called Tega, was used in an affective tutoring system. This experiment featured a game, designed for a tablet, that aimed at teaching children a second language with the collaboration of a social robot in order to determine each student's affective state throughout the tutoring session and personalize each session according to that information. The authors measured the children's valence and engagement with a facial expression analysis system and used the extracted information to improve the robot's algorithm. The results of this two-month experiment showed that the children acquired new knowledge, and the robot managed to adapt the learning process for each student.

Finally, the use of robotics in project-based education needs to be noted as well. Several works in the literature [35–38] described the application of educational robots in project-based teaching, and the results of the works were positive.

### 3. System Modules

The research methodology that was used in this study is described as follows. The first step was an extensive literature review of the related scientific literature. The second step involved the design of the robotics system with NAO based on the input of the literature review and the consideration of emotion recognition techniques. The third step was the implementation of the presented system. The application was used in real-life circumstances and was evaluated.

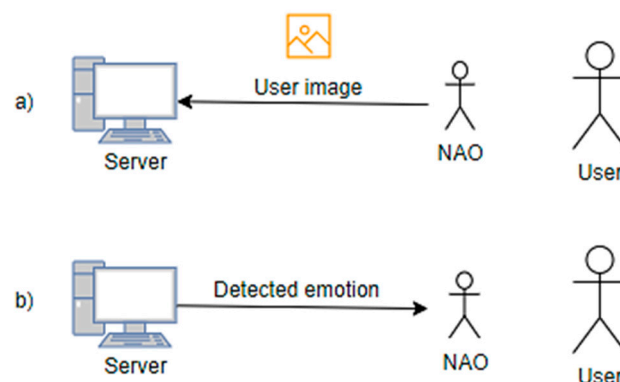
The modules of the developed system consist of a face attribute analyzer that recognizes emotions, a client–server socket module for image transferring, and the NAO robot as a conversation agent.

#### 3.1. General Functionality

In order to create an optimal prototype to be used in real-life scenarios, the functionality was divided into five steps.

- (1) The humanoid robot NAO is tasked with greeting the user and explaining the game rules, including that the final outcome can be affected by the answers, but the user can always change an answer if NAO notices negative emotions. NAO also explains how the emotion system works, asking them to assume a facial expression of either happiness, anger, fear, disgust, surprise, sadness, or neutrality each time the result of the answer is announced.
- (2) It was decided that the robot should wait 1 s for the user to assume their choice of expression and then capture one image via the web cameras located on its forehead.
- (3) The robot sends the image to the server, where the emotion recognition process is set to start, and waits for a result.
- (4) The server saves the image and uses it on a face attribute analyzer in order to extract the most dominant emotion.
- (5) The result is transferred to the robot, which acts accordingly, e.g., moving on to the next question if the user is happy or allowing for a new answer if the user is sad.

Figure 1 shows the flow of data, with the user image as the input to the face attribute analyzer and the detected emotion as the output produced by the system and transferred to the robot.



**Figure 1.** (a) NAO captures an image of the user and sends it to the server. (b) The server replies with the extracted emotion.

#### 3.2. Face Attribute Analyzer

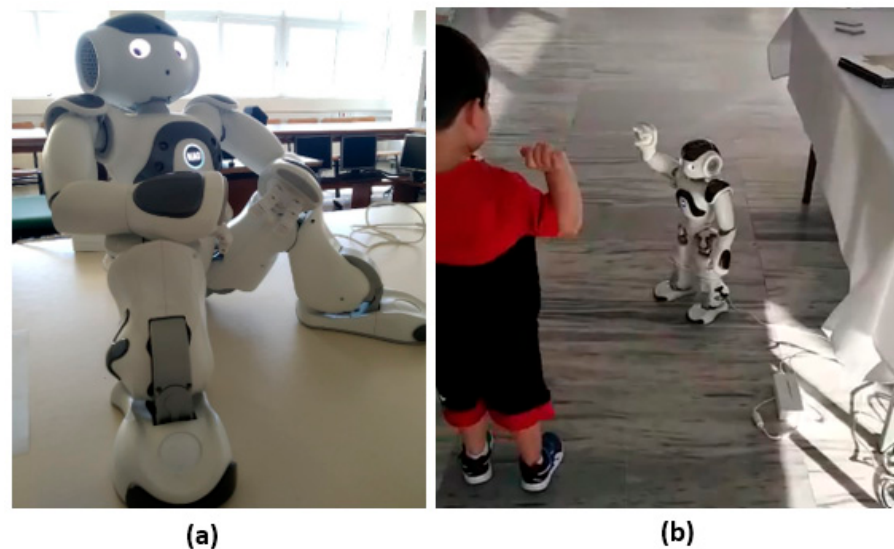
The chosen face attribute analyzer was DeepFace [14], a deep learning facial recognition framework for Python, which is able to identify age, gender, emotion, and race in digital images. The three major factors that resulted in the use of this facial recognition

system were the high accuracy rates, sometimes making DeepFace more successful than humans (97.35% on the Labeled Faces in the Wild (LFW) data set, contrary to 97.5% from human beings) [39]; the ability to recognize the six basic emotions, anger, fear, happiness, sadness, surprise, disgust (FACS Ekman 1978), and neutrality; and the overall fast execution time. It is also lightweight and easy to install.

### 3.3. NAO and Choregraphe

NAO is a humanoid robot created by SoftBank Robotics. Although small (58 cm in height), it features 25 degrees of freedom, allowing it to move and adapt to its surroundings, and various sensors, microphones, and cameras used for speech, object, and face recognition.

Choregraphe, a platform dedicated to creating software for SoftBank Robotics' humanoid robots, was used for creating NAO's behavior. This platform features a graphical user interface (GUI) and allows the use of "boxes", which are Python scripts with specific functionalities. Most of the boxes are customizable, and the user can create new Python scripts in order to use a specific feature. The boxes can connect with each other to send and receive data. Figure 2 shows the project behavior that was tested under real conditions.



**Figure 2.** (a) The humanoid robot NAO; (b) NAO greeting a child.

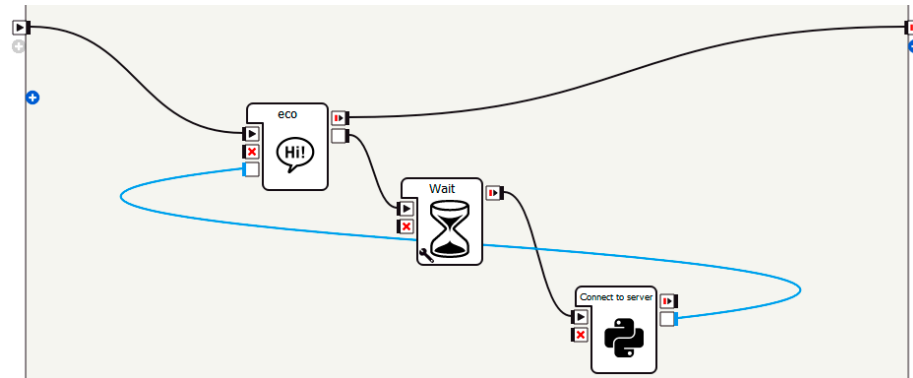
NAO's behavior schematic consists of three boxes (Figure 3). The first one is a dialog box, which allows the user to write a series of rules, creating scenario dialogues. These scenarios are triggered by either user input or proposals. The user input is any human input that matches with the rule, while the proposal does not require human input in order to be triggered. This dialog box contained both user rules and proposals, and the topic referred to the first and last steps of the proposed general functionality: greeting the user, explaining the task, and acting accordingly to the user's detected emotion.

The second box was available directly from Choregraphe's library. Its functionality was to create a time-based delay based on a given parameter. This box was used for generating a one-second delay to ensure the user assumed their chosen expression.

The third box contained a Python script that was used for the second, third, and fifth steps: the capturing of the image, its transmission to the server, and the receipt of the dominant emotion. ALVideoDevice—NAO's module responsible for providing images from video sources—handled the snapshot. The "send and receive processes" were handled by Python sockets, following a client–server model, with the client being the NAO robot and the server being a computer. The client was tasked with sending the image and retrieving the result. The server managed the fourth step: saving the image and executing the face attribute analyzer.



The aforementioned behavior can be executed in a loop or modified accordingly by changing or adding rules inside the dialog box. The waiting time can also be modified or omitted.



**Figure 3.** Modules of NAO's behavioral patterns.

## 4. System Architecture

### 4.1. Image Capture and Sockets

To capture the image via NAO's web cameras, we utilized the `ALVideoDevice` module. We needed to subscribe to the proxy by setting the resolution, the color space, and the minimum frames per second. After the subscribe function was called, we could use the module's built-in methods to obtain the image from the camera. This was done by requesting the image and then releasing the buffer.

The client socket code followed the TCP standard, which allows for packet transferring across the internet in a reliable way, meaning that the packets arrive in order and without errors.

We took the following steps: specifying the host IP address and the port number, establishing connection with the server, sending the buffer size of the image, sending the image data, and waiting for a response. All of the sent data were serialized by the `Pickle Python` module [10] in order to convert them to bytes. The images in NAO are returned in the form of a container object and have 12 fields (0 to 11). Field number six corresponds to the image buffer, which is a binary array of size  $\text{height} \times \text{width} \times \text{number of layers}$  containing image data [40]. At the end of that buffer, we added a string that signified the end of the data. This was used by the server to detect the last packet sent. The response consisted of a string that described the detected emotion (extracted by the analyzer). The emotion was stored in a variable and transferred over to the dialog box by utilizing box variables and the connections between the boxes.

The steps followed for the server code were as follows:

- (1) Specify the IP address and port number;
- (2) Bind the socket to the specified IP and port;
- (3) Listen for new connections;
- (4) The server establishes a connection with the client and initializes the data transfer. The data are sent in chunks until the added string that signifies the end of data is met. Each data chunk is appended to an array, which, by the end of the data transfer, holds all of the image data. Algorithm 1 shows the data reception.
- (5) The `Pickle` module is used to reconstruct the image using the data stored in the array. The reconstructed image is saved on the server.
- (6) The emotion analyzer is called afterwards, utilizing the image as input in order to return the dominant emotion. When the emotion is extracted, the server sends the result to the client and the connection is terminated.

**Algorithm 1.** Receive data from client.

---

**Input:** A socket connection  
**Output:** The image data in bytes  
**Initialization of variables:** Assign an empty byte array to variable final\_data

```

1  while (true) do
2      try
3          img_data ← receive packet(data)
4      catch
5          exit
6          binary_array ← img_data
7          if img_data endswith "[END OF IMAGE]" then
8              binary_array ← replace "[END OF IMAGE]" with empty string
9              final_data ← binary_array
10             exit
11         else
12             final_data ← final_data + binary_array
13     end

```

---

## 4.2. NAO and Choregraphe

NAO's ALDialog module allows the creation of human-like dialogs by utilizing a list of rules written by the user. These rules can be grouped in topics, and they should be categorized in appropriate ways. The two types of rules, user and proposals, are either triggered by user input, which links to a possible robot output, or by the robot itself with topic progression functions [17]. The user rules in the proposed method are triggered by an event bound to an input variable located on the box, which stores the detected emotion, passed on by the Python script. In Figure 4, the input variable is called "user\_emotion", and the event (symbolized as e:user\_emotion) triggers every time an emotion is recognized. There is a separate rule for each of the emotions, which allows the robot to react with different and unique reactions for all of the emotions.

```

u: ("e:user_emotion $user_emotion=="sad") Why are you sad? Wanna hear a joke?
  u1: (~positive) Why was six afraid of seven? Cause seven eight nine!
$Output=1
  u2: ("e:user_emotion $user_emotion=="happy") You're happy. My joke
worked, yay! $onStopped=1
  u2: ("e:user_emotion $user_emotion=="sad") You're still sad. I'll try
harder next time. $onStopped=1
  u2: ("e:user_emotion $user_emotion=="fear") Don't be afraid. I think
i made things worse. $onStopped=1
  u2: ("e:user_emotion $user_emotion=="disgust") Alright, I don't have
a sense of humor, but don't be disgusted! $onStopped=1
  u2: ("e:user_emotion $user_emotion=="neutral") Hmm. That's a bit
better. $onStopped=1
  u2: ("e:user_emotion $user_emotion=="surprised") I'm surprised that
you are surprised by my joke. $onStopped=1
  u1: (~negative) I'm sure I can help you some other way. $onStopped=1

```

**Figure 4.** Example QiChat dialogue used during initial testing.

A dialog box can remain active throughout the entire conversation, even when the behavior shifts to the other boxes. This functionality allows the creation of subrules, which resemble the user rules with the exception that they cannot be triggered without the main user rule being active. This feature can be used for more in-depth conversations with the robot as they can still be activated by an emotion without triggering other main user rules or having to execute the behavior multiple times.

## 5. Evaluation Results

The project was used and tested under real-life circumstances during an environmental festival, which aimed to create environmental awareness and inform people about sustainable transportation and smart cities while showcasing new, eco-friendly technology.

The sample was 50 attendees of the festival who were chosen randomly and were divided into two groups of 25 members with similar demographical characteristics, namely, Group A, who used a web application through laptops, and Group B, who used NAO. In particular, the sample consisted of 28 (56%) male and 22 (44%) female students, having ages ranging from 11 to 14 years old. All participants lived in urban centers and had experience with using a computer. It should be noted that the data used for the experiment were collected and held anonymously.

NAO was tasked with sharing tips and information about recycling with the attendees through the developed quiz, recognizing users' emotions while interacting with them and providing tailored support to user's emotions when a wrong answer was given. On the other hand, in the web application used, the same quiz was developed, but a simple second chance was given in every incorrect answer, without tailored support or emotion recognition.

After completing the interaction with the systems, the users were asked to rate, using five-point scale (1: very low, 2: low, 3: medium, 4: high, 5: very high), whether they enjoyed the interaction with the system and whether the system helped them get acquainted with environmental issues. According to the answers, a two-sample t-test was performed to compare the benefits of using NAO against a conventional application. Table 1 illustrates the t-test results. As shown, there was a significant difference in the degree of enjoying the interaction between NAO ( $M = 4.8$ ,  $SD = 1.292$ ) and the conventional application ( $M = 3.28$ ,  $SD = 0.843$ );  $t(48) = 8.117$ ,  $p = 1.47 \times 10^{-10}$ . Moreover, there was a significant difference in the degree of getting acquainted with environmental issues between NAO ( $M = 4.4$ ,  $SD = 0.707$ ) and the conventional application ( $M = 3.12$ ,  $SD = 0.726$ );  $t(48) = 6.316$ ,  $p = 8.23 \times 10^{-8}$ .

**Table 1.** T-test results: Comparing NAO with conventional application.

	Did You Enjoy Interacting with the System?		Did the System Help You Get Acquainted with Environmental Issues?	
	Group A	Group B	Group A	Group B
Mean	4.8	3.28	4.4	3.12
Variance	0.167	0.71	0.5	0.527
Observations	25	25	25	25
Pooled Variance	0.438		0.513	
Hypothesized Mean Difference	0		0	
df	48		48	
t Stat	8.117		6.316	
P(T <= t) two-tail	$1.47 \times 10^{-10}$		$8.23 \times 10^{-8}$	
t Critical two-tail	2.011		2.011	

The findings revealed that the participants who interacted with NAO greatly enjoyed their conversations with it. Their experience with NAO was characterized as entertaining and engaging. This positive feedback may be due to NAO's appearance and behavior, simulating human reactions. Moreover, interacting with a humanoid robot was a challenging experience for the users, attracting their interest in comparison to the use of a conventional application. The participants emphasized that NAO's reactions and movements, such as greeting, moving hands, etc., made them feel more comfortable with it. Furthermore, the emotion recognition and tailored feedback based on it made conversations with NAO more personalized, improving user engagement and retention. This personalized verbal communication motivated the participants and encouraged them to be more communicative. As a result, the participants who used NAO noted that they were aided in getting acquainted with environmental topics more than those who interacted with the conventional application where there was no tailored feedback. In summary, the evaluation results showed



the potential of NAO technology, especially when combined with tailored reactions to users' emotions.

In order to evaluate the emotion recognition functionality, the participants who used NAO were asked to adopt a facial expression of their choice in front of NAO's cameras. NAO responded by saying something related to the expression. The participants were then be asked to state the feeling they wanted to express. The results of the emotion evaluation were considered correct if the extracted emotion matched the intended emotion or if it was related to the given facial expression, even if it was not the user's intended emotion. This was decided due to the fact that some of the participants stated that they wanted to express a specific emotion but then eventually assumed a different expression, as NAO's appearance generally encourages positive emotions, or because they assumed it very briefly. Tables 2 and 3 and Figure 5 illustrate the results of the emotion evaluation.

**Table 2.** Results of emotion evaluation.

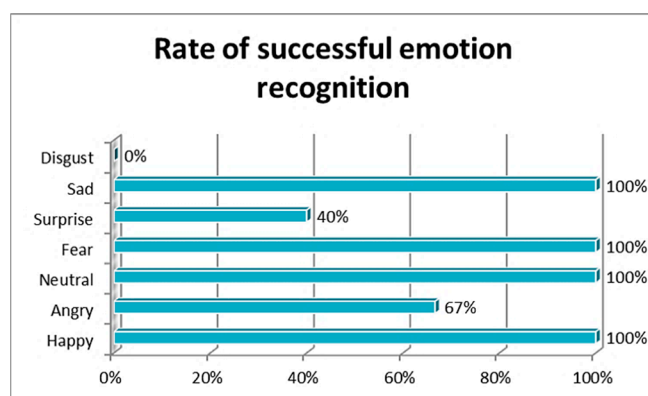
	Human Input	Emotion Analyzer Output
1	Happy	Happy
2	Angry	Angry
3	Neutral	Neutral
4	Happy	Happy
5	Fear	Fear
6	Surprise	Surprise
7	Neutral	Neutral
8	Happy	Happy
9	Fear	Disgust <sup>1</sup>
10	Surprise	Fear <sup>1</sup>
11	Neutral	Neutral
12	Surprise	Angry
13	Sad	Sad
14	Neutral	Neutral
15	Sad	Sad
16	Happy	Happy
17	Angry	Neutral
18	Disgust	Sad
19	Happy	Happy
20	Sad	Sad
21	Fear	Disgust <sup>1</sup>
22	Fear	Fear
23	Surprise	Fear
24	Angry	Angry
25	Surprise	Sad

In emotion analyzer output column, the correct results have been marked with green color, whereas the unmatched ones have been marked with red color. <sup>1</sup> These emotions were measured correctly as participants' expressions could be interpreted with the results returned by the analyzer. These cases were also discussed with the participants.

Overall, the test results were very encouraging. Out of 25 answers, 5 of them were found incorrect. The expressions that were shown to be the most confusing were anger (67% successful emotion recognition), disgust (unsuccessful recognition), and surprise (40% successful emotion recognition); meanwhile, sadness, fear, neutral, and happy emotions were recognized successfully. Light sources played a major role, and some of the participants' pictures often appeared too dark or too bright, depending on the sun, as the assessment was conducted in a room with large windows. In these cases, it was decided that the evaluation should be run again with a new picture from the same participant.

**Table 3.** Confusion matrix of emotion prediction.

		Predicted							
		Happy	Angry	Neutral	Fear	Surprise	Sad	Disgust	
Actual	Happy	5	0	0	0	0	0	0	5
	Angry	0	2	1	0	0	0	0	3
	Neutral	0	0	4	0	0	0	0	4
	Fear	0	0	0	4	0	0	0	4
	Surprise	0	1	0	1	2	1	0	5
	Sad	0	0	0	0	0	3	0	3
	Disgust	0	0	0	0	0	1	0	1
	5	3	5	5	2	5	0	N = 25	

**Figure 5.** Rate of successful emotion recognition.

Regarding the execution/response speed, the entire conversation with the robot was equivalent to a human–human conversation. Excluding the initial run, which required some additional time due to setting up, the execution and response times were very brief, and none of the participants showed any negative signs during the waiting time, which was, on average, 10 s for the entire conversation.

## 6. Conclusions, Limitations, and Future Work

The NAO robot seems to be very powerful in the fields of entertainment and education, especially with young children, as this was observed during the festival. Its humanoid looks and prebuilt functionality for conversations make it an optimal candidate for human–robot interaction, as people were willing to partake in conversations with it. Combined with the DeepFace API, it can achieve a fast and reliable way of analyzing people’s emotions.

In this paper, we designed a reliable method for data transferring between the NAO robot and a computer, and we used that data with an external face attribute analyzer in order to detect human emotions. With the given results, we are able to generate different responses from NAO that befitted each situation. In particular, the NAO robot was adopted as an experimental component for natural conversations with constructed dialogues and actions to interact with humans and was used experimentally under real-life circumstances in order to teach people about recycling in a fun and interactive way. This experiment yielded favorable results both in terms of emotion recognition as well as entertaining and teaching people.

### 6.1. Limitations

When running the project in the real robot, it was observed that the image data appeared to be truncated, thereby making image reconstruction on the server side impossible. While this was not the case with the Choregraphe’s simulator, a workaround was designed by attaching a string that signified the end of the image data at the end of the transferred data, which allowed the server to receive the entire image. Moreover, the time needed for

the first setup should be taken into consideration when deploying the project in real-life scenarios, along with the available light sources and camera angles, in order to achieve optimal results.

Another limitation was the consideration of users' self-perception concerning their acquaintance with environmental issues in the evaluation phase.

## 6.2. Future Work

By focusing on the easier use and deployment of the project as well as optimization and better results, some adjustments and improvements can be made. Giving the robot the option of free movement without having to be near the server would allow for faster deployment and eliminate the need for a proper environment to be set up each time the robot is used. This can be achieved with a Jetson Nano microcomputer strapped on the robot's back that can run the socket and the emotion analyzer code. Furthermore, combining facial expressions and speech during the analysis can lead to higher success percentages as speech can provide even more information about the person's current emotional state. This can go in conjunction with the use of different models or neural networks as they can be tailored to focus on specific characteristics, support more emotions, or allow for speech emotion analysis in the user's native language.

Finally, part of our future plans is a more in-depth evaluation that will be based on more collected users' data and will include factor analysis.

**Author Contributions:** Conceptualization, I.A.V., C.T. and A.K.; methodology, I.A.V., C.T. and A.K.; software, I.A.V., C.T. and A.K.; validation, I.A.V., C.T. and A.K.; formal analysis, I.A.V., C.T. and A.K.; investigation, I.A.V., C.T. and A.K.; resources, I.A.V., C.T. and A.K.; data curation, I.A.V., C.T. and A.K.; writing—original draft preparation, I.A.V., C.T. and A.K.; writing—review and editing, I.A.V., C.T. and A.K.; visualization, I.A.V., C.T. and A.K.; supervision, M.F. and C.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data used to support the findings of this study have not been made available because they contain information that could compromise research participant privacy/consent.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Troussas, C.; Espinosa, K.J.; Virvou, M. Affect Recognition through Facebook for Effective Group Profiling Towards Personalized Instruction. *Inform. Educ.* **2016**, *15*, 147–161. [\[CrossRef\]](#)
2. Krouska, A.; Troussas, C.; Virvou, M. Deep Learning for Twitter Sentiment Analysis: The Effect of Pre-trained Word Embedding. In *Machine Learning Paradigms. Learning and Analytics in Intelligent Systems*; Tsihrintzis, G., Jain, L., Eds.; Springer: Cham, Switzerland, 2020; Volume 18. [\[CrossRef\]](#)
3. Pei, E.; Zhao, Y.; Oveneke, M.C.; Jiang, D.; Sahli, H. A Bayesian Filtering Framework for Continuous Affect Recognition from Facial Images. *IEEE Trans. Multimed.* **2022**, *1*. [\[CrossRef\]](#)
4. Troussas, C.; Krouska, A.; Virvou, M. Trends on Sentiment Analysis over Social Networks: Pre-processing Ramifications, Stand-Alone Classifiers and Ensemble Averaging. In *Machine Learning Paradigms. Intelligent Systems Reference Library*; Tsihrintzis, G., Sotiropoulos, D., Jain, L., Eds.; Springer: Cham, Switzerland, 2019; Volume 149. [\[CrossRef\]](#)
5. Troussas, C.; Krouska, A.; Virvou, M. A Multicriteria Framework for Assessing Sentiment Analysis in Social and Digital Learning: Software Review. In Proceedings of the 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), Zakynthos, Greece, 23–25 July 2018; pp. 1–7. [\[CrossRef\]](#)
6. Caruelle, D.; Shams, P.; Gustafsson, A.; Lervik-Olsen, L. Affective Computing in Marketing: Practical Implications and Research Opportunities Afforded by Emotionally Intelligent Machines. *Mark. Lett.* **2022**, *33*, 163–169. [\[CrossRef\]](#)
7. Andrej, L.; Panagiotis, B.; Madga, H.-A. Affective computing and medical informatics: State of the art in emotion-aware medical applications. *Stud. Health Technol. Inform.* **2008**, *136*, 517–522. [\[CrossRef\]](#)

8. Stein, G.; Ledeczi, A. Enabling Collaborative Distance Robotics Education for Novice Programmers. In Proceedings of the 2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), St Louis, MO, USA, 10–13 October 2021; pp. 1–5. [CrossRef]
9. Ververi, C.; Koufou, T.; Moutzouris, A.; Andreou, L.-V. Introducing Robotics to an English for Academic Purposes Curriculum in Higher Education: The Student Experience. In Proceedings of the 2020 IEEE Global Engineering Education Conference (EDUCON), Porto, Portugal, 27–30 April 2020; pp. 20–21. [CrossRef]
10. Tengler, K.; Kastner-Hauler, O.; Sabitzer, B. A Robotics-based Learning Environment Supporting Computational Thinking Skills—Design and Development. In Proceedings of the 2021 IEEE Frontiers in Education Conference (FIE), Lincoln, NE, USA, 13–16 October 2021; pp. 1–6. [CrossRef]
11. Cavedini, P.; Bertagnolli, S.D.C.; Peres, A.; Oliva, R.S.; Locatelli, E.L.; Caetano, S.V.N. Educational Robotics and Physical Education: Body and movement in learning laterality in Early Childhood Education. In Proceedings of the 2021 International Symposium on Computers in Education (SIIE), Malaga, Spain, 23–24 September 2021; pp. 1–6. [CrossRef]
12. NAO Robot. Available online: <https://www.softbankrobotics.com/emea/en/nao> (accessed on 1 April 2022).
13. SoftBank Robotics NAO Documentation: ALMood. Available online: <https://doc.aldebaran.com/2-5/naoqi/core/almoood.html> (accessed on 1 April 2022).
14. DEEPFACE, Face Recognition and Facial Attribute Analysis Library for Python. Available online: <https://github.com/serengil/deepface> (accessed on 1 April 2022).
15. Python PICKLE Module. Available online: <https://github.com/python/cpython/blob/3.10/Lib/pickle.py> (accessed on 1 April 2022).
16. Choregraphe Suite. Available online: [https://doc.aldebaran.com/2-8/software/choregraphe/choregraphe\\_overview.html](https://doc.aldebaran.com/2-8/software/choregraphe/choregraphe_overview.html) (accessed on 1 April 2022).
17. SoftBank Robotics NAO Documentation: QiChat. Available online: <https://doc.aldebaran.com/2-1/naoqi/audio/dialog/dialog.html#dialog-concepts> (accessed on 1 April 2022).
18. She, T.; Ren, F. Enhance the Language Ability of Humanoid Robot NAO through Deep Learning to Interact with Autistic Children. *Electronics* **2021**, *10*, 2393. [CrossRef]
19. Ismail, S.B.; Shamsuddin, S.; Yussof, H.; Hashim, H.; Bahari, S.; Jaafar, A.; Zahari, I. Face detection technique of Humanoid Robot NAO for application in robotic assistive therapy. In Proceedings of the 2011 IEEE International Conference on Control System, Computing and Engineering, Penang, Malaysia, 25–27 November 2011; pp. 517–521. [CrossRef]
20. Torta, E.; Werner, F.; Johnson, D.O.; Juola, J.F.; Cuijpers, R.H.; Bazzani, M.; Oberzaucher, J.; Lemberger, J.; Lewy, H.; Bregman, J. Evaluation of a Small Socially-Assistive Humanoid Robot in Intelligent Homes for the Care of the Elderly. *J. Intell. Robot. Syst.* **2014**, *76*, 57–71. [CrossRef]
21. Scassellati, B.; Boccanfuso, L.; Huang, C.-M.; Mademtzi, M.; Qin, M.; Salomons, N.; Ventola, P.; Shic, F. Improving social skills in children with ASD using a long-term, in-home social robot. *Sci. Robot.* **2018**, *3*, eaat7544. [CrossRef] [PubMed]
22. Ramis, S.; Buades, J.M.; Perales, F.J. Using a Social Robot to Evaluate Facial Expressions in the Wild. *Sensors* **2020**, *20*, 6716. [CrossRef] [PubMed]
23. Lopez-Rincon, A. Emotion recognition using facial expressions in children using the NAO Robot. In Proceedings of the International Conference on Electronics, Communications and Computers (CONIELECOMP), Cholula, Mexico, 27 February–1 March 2019; pp. 146–153. [CrossRef]
24. Hu, M. Facial Emotional Recognition with Deep Learning on Pepper Robot. Bachelor’s Thesis, Vaasan Ammattikorkeakoulu University of Applied Sciences, Vasa, Finland, 2019.
25. Onder, T.; Fatma, G.; Duygun, E.B.; Hatice, K. An Emotion Analysis Algorithm and Implementation to NAO Humanoid Robot. *Eurasia Proc. Sci. Technol. Eng. Math. EPSTEM* **2017**, *1*, 316–330.
26. Filippini, C.; Perpetuini, D.; Cardone, D.; Merla, A. Improving Human–Robot Interaction by Enhancing NAO Robot Awareness of Human Facial Expression. *Sensors* **2021**, *21*, 6438. [CrossRef] [PubMed]
27. SoftBank Robotics NAO Documentation: NAOqi. Available online: <https://doc.aldebaran.com/2-5/naoqi/core/index.html> (accessed on 1 April 2022).
28. Castellano, G.; De Carolis, B.; Macchiarulo, N.; Rossano, V. Learning waste Recycling by playing with a Social Robot. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019; pp. 3805–3810. [CrossRef]
29. PEPPER Robot. Available online: <https://www.softbankrobotics.com/emea/en/pepper> (accessed on 1 April 2022).
30. Vrochidou, E.; Najoua, A.; Lytridis, C.; Salonidis, M.; Ferelis, V.; Papakostas, G.A. Social Robot NAO as a Self-Regulating Didactic Mediator: A Case Study of Teaching/Learning Numeracy. In Proceedings of the 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 13–15 September 2018; pp. 1–5. [CrossRef]
31. Ozkul, A.; Kose, H.; Yorganci, R.; Ince, G. Robostar: An interaction game with humanoid robots for learning sign language. In Proceedings of the 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO), Bali, Indonesia, 5–10 December 2014; pp. 522–527. [CrossRef]
32. Ali, S.; Moroso, T.; Breazeal, C. Can Children Learn Creativity from a Social Robot? In Proceedings of the 2019 on Creativity and Cognition (C&C ‘19). Association for Computing Machinery, New York, NY, USA, 23–26 June 2019; pp. 359–368. [CrossRef]

33. Miskam, M.A.; Shamsuddin, S.; Samat, M.R.A.; Yussof, H.; Ainudin, H.A.; Omar, A.R. Humanoid robot NAO as a teaching tool of emotion recognition for children with autism using the Android app. In Proceedings of the 2014 International Symposium on Micro-NanoMechatronics and Human Science (MHS), Nagoya, Japan, 10–12 November 2014; pp. 1–5. [[CrossRef](#)]
34. Gordon, G.; Spaulding, S.; Westlund, J.K.; Lee, J.J.; Plummer, L.; Martinez, M.; Das, M.; Breazeal, C. Affective personalization of a social robot tutor for children’s second language skills. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI’16), 12–17 February 2016; pp. 3951–3957.
35. Štuikys, V.; Burbaite, R.; Damaševičius, R. Teaching of Computer Science Topics Using Meta-Programming-Based GLOs and LEGO Robots. *Inform. Educ.* **2013**, *12*, 125–142. [[CrossRef](#)]
36. Damaševičius, R.; Narbutaitė, L.; Plauska, I.; Blažauskas, T. Advances in the Use of Educational Robots in Project-Based Teaching. *TEM J.* **2017**, *6*, 342–348.
37. Dederichs-Koch, A.; Zwierns, U. Project-based learning unit: Kinematics and object grasping in humanoid robotics. In Proceedings of the 2015 16th International Conference on Research and Education in Mechatronics (REM), Bochum, Germany, 18–20 November 2015; pp. 216–220. [[CrossRef](#)]
38. Karaman, S.; Anders, A.; Boulet, M.; Connor, J.; Gregson, K.; Guerra, W.; Guldner, O.; Mohamoud, M.; Plancher, B.; Shin, R. Project-based, collaborative, algorithmic robotics for high school students: Programming self-driving race cars at MIT. In Proceedings of the 2017 IEEE Integrated STEM Education Conference (ISEC), Princeton, NJ, USA, 11 March 2017; pp. 195–203. [[CrossRef](#)]
39. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708. [[CrossRef](#)]
40. SoftBank Robotics NAO Documentation: ALVideoDevice. Available online: <http://doc.aldebaran.com/2-1/naoqi/vision/alvideodevice-api.html> (accessed on 1 April 2022).