Predicting Anthropomorphic Attitudes towards Robots based on Facial Expressions

Valentin Oliver Loftsson
SCIPER: 308099

ABSTRACT

This study investigates anthropomorphism towards robots and whether it can be predicted from facial expressions. We present an online experiment that was conducted for two weeks. Participants watched three video clips of robots of varying degrees of human likeness while their face was recorded through their web camera. Meanwhile, their facial expressions were predicted in real-time using a well-performing facial action classifier. Following each video, the anthropomorphic response was measured through a quick survey. The data was pre-processed and aggregated to produce samples to train and validate a machine learning model that predicts the anthropomorphic response from the facial expressions. The least absolute shrinkage and selection operator (LASSO) method was used with cross-validation and the coefficient of determination is reported for each model. Moreover, detailed data analysis was performed. In the future, it is hoped that the approach will be refined as more data becomes available.

INTRODUCTION

The discipline of human-robot interaction (HRI) has gathered momentum in the past decades due to rapid advancements in computer science and artificial intelligence. Research in HRI has mainly focused on human responses and attitudes towards robots (Dautenhahn, 2014). Less has been investigated on how robots can, without verbal instructions, infer human feelings from observations, and adapt their behavior accordingly. Such an inference feedback model enables further sophistication of the decision-making mechanism of the robot, encouraging healthy rather than overly dependent behaviors and leading to more pleasant interactions. For instance, a robot with such capabilities might decide to retire from a person's presence if it infers that the person would like to be left alone. This study investigates the feasibility of developing a model that can predict the degree of anthropomorphism towards a robot from facial expressions. Anthropomorphism, as defined by Bartneck, et al., is "the attribution of a human form, human characteristics, or human behavior to nonhuman things such as robots, computers, and animals" (Bartneck, Kulic, Croft, & Zoghbi, 2008, page 74).

Undeniably, robots of every kind will become more widespread in years to come and human-robot coexistence entails challenges that must be addressed. Robot developers of today face the challenge of creating robots that meet human expectations and accomplish social goals (Kiesler & Goetz, 2002). Naturally, humans are most comfortable interacting with humans, and it has been suggested that anthropomorphism towards robots is an indication of acceptability, along with other factors (Eyssel, Kuchenbrandt, & Bobinger, 2011; Bartneck, et al., 2008). To measure anthropomorphic attitudes, studies have often carried out questionnaires following some interaction with or observation of robots (MacDorman, 2006; Powers & Sara, 2006; Kiesler & Goetz, 2002). These questionnaires use different items and scales which makes it hard to compare studies. Bartneck, et al. (2008) performed a literature

review and proposed standardized “Godspeed” questionnaires for measuring anthropomorphism and other key factors in HRI.

People communicate attitudes and feelings via language, facial expressions, and whole-body gestures. Van den Stock, Righart, & de Gelder (2007) found that whole-body expressions are important in communication and can help with emotion-recognition, particularly when facial expressions are ambiguous. Moreover, Gunes & Piccardi (2007) and Caridakis, et al. (2007) demonstrated that multimodal approaches to emotion classification, that integrate speech, face and body expressions, outperform unimodal models that involve only one of these channels. Nevertheless, the face is the richest non-verbal channel of communication and is still a good emotion indicator by itself. The AFFDEX SDK is one example of a unimodal emotion classifier based solely on facial expressions (McDuff, et al., 2016). Relatively few studies have been carried out that exploit such tools to derive facial expressions that can, in turn, be used to improve robot behavior without the need for verbal instructions. Furthermore, no studies have focused on anthropomorphism specifically.

In this study, participants go through an experiment on a tailor-made web site (Loftsson & Griesser, 2020). They watch three video clips of robots while their face is recorded, and facial expressions are detected in real-time using the AFFDEX classifier. Following each video clip, the anthropomorphic response is measured by the Godspeed questionnaire. These data are then used to train and validate a machine learning model that predicts the anthropomorphic response from the facial expressions.

How well can anthropomorphic attitudes towards robots be predicted from facial expressions? It is expected that there exist strong relationships between the degree to which people anthropomorphize robots and their facial expressions. This proposition is tested via correlation analysis and model evaluation techniques.

METHOD

Participants

The sample consists of volunteer participants aged 17-80 of both genders. Participants are required to have an intermediate ability in the English language, as instructions and content are in English. Participation is anonymous and no video recordings or personal information aside from age and gender are stored.

Materials and measures

The study involves three short video clips of robots of varying degrees of human-likeness and sophistication. The least humanlike robot is the robot arm (Figure 1) (AUBO Robotics, 2017), the second humanlike robot is Pepper the Robot (Figure 2) (MobileSyrup, 2016), and the most human-like robot is the social humanoid Sophia (Figure 3) (funtime 247, 2017). The playback ranges were carefully selected to capture the most significant clips of each video. The length of each clip is 1:00-1:20 minute.

Figure 1: Robot arm pouring tea

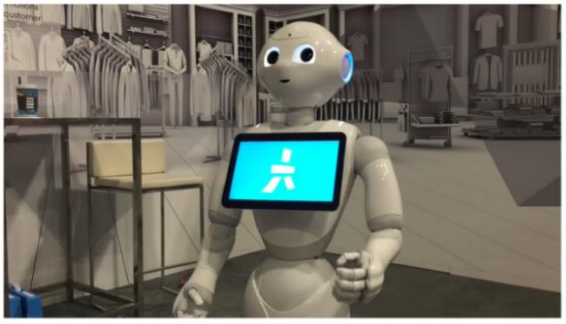

Figure 2: Pepper the Robot

Figure 3: Sophia the Robot

Facial expressions are extracted from the camera stream using a facial action causal classifier provided by Affectiva's AFFDEX JavaScript software development kit (SDK) (McDuff, et al., 2016). By "causal" classifier it is meant that video streams are processed rather than static images to obtain better classification accuracy. The SDK classifies each face into several facial actions (AUs), also known as the building blocks of facial expressions (see Figure 4). Each action is given a score from 0 (absent) to 100 (present). These scores represent the predictor variables, as illustrated in Figure 5.

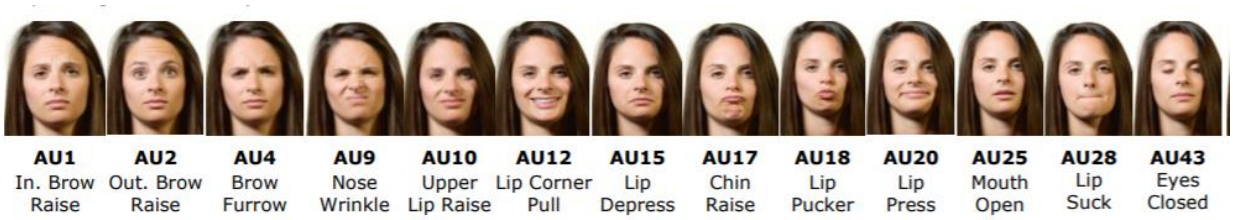

Figure 4: Some of the facial actions predicted by the AFFDEX SDK. Source: McDuff, et al. (2016).

	Nose wrinkle (1)	Lip corner pull (2)	Lip depress (3)	Chin raise (4)	...
$\mathbf{s_1}$	s_{11}	s_{12}	s_{13}	s_{14}	...
$\mathbf{s_2}$	s_{21}	s_{22}	s_{23}	s_{24}	...
$\mathbf{s_3}$	s_{31}	s_{32}	s_{33}	s_{34}	...
...	...	...	...	...	...

Figure 5 : Simple illustration of the feature matrix representation of the facial actions of participants. Vector $\mathbf{s_i}$ is the facial action score vector for participant **i** and element $\mathbf{s_{ij}}$ of this vector represents the score for facial action **j**.

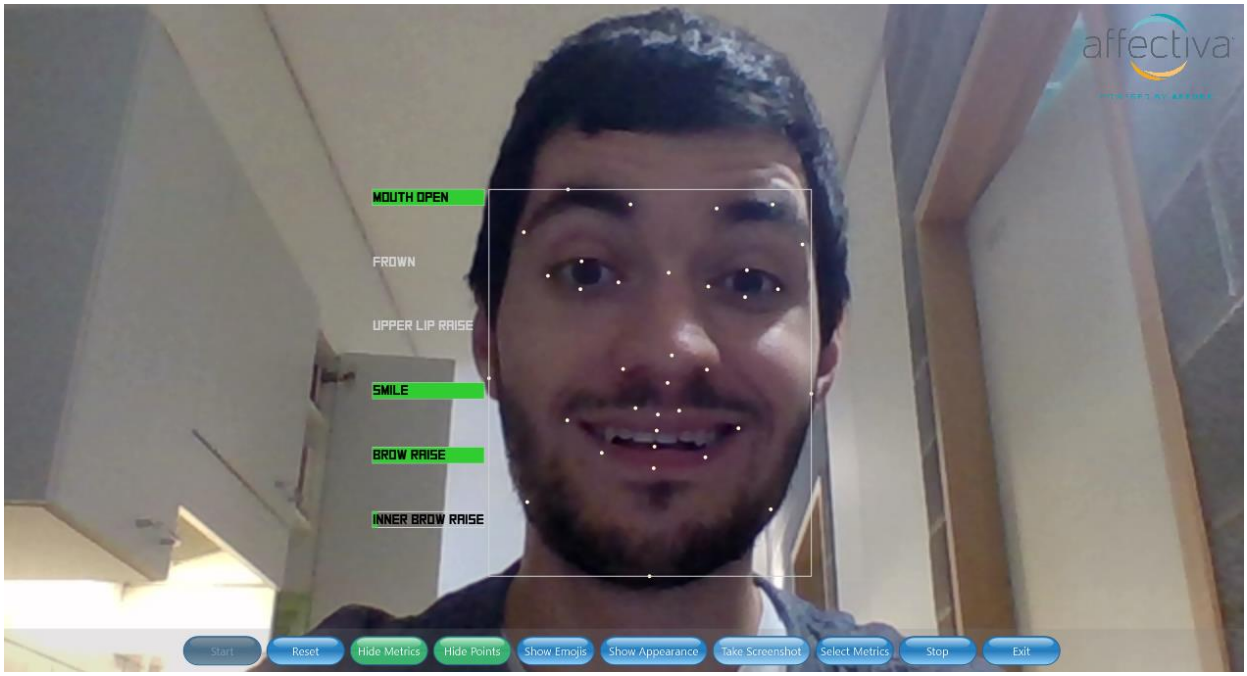

Figure 6: Demonstrating the use of the AFFDEX SDK in the demo app *AffdexMe*. Facial feature points are rendered on the face and the ranks of 6 pre-selected facial actions are displayed to the left.

Affectiva's JavaScript SDK was selected for the current study because it is widely used and easy to use with a client-side web application. Moreover, the underlying model performs better than the previously published baselines (Senechal, McDuff, & el Kaliouby, 2015). The main reasons for the system's success are: 1) the classifiers were trained on the world's largest annotated dataset of facial expressions, the size of which was achieved by the efficient active learning approach to the hand-labeling process, and 2) the Nyström kernel approximation method was used in training the classifiers to achieve high accuracy and performance.

The participant's impression of each robot is assessed by a short survey immediately following the respective video clip. The survey is based on the standardized Godspeed questionnaires for HRI (Bartneck, et al., 2008). The questionnaires use a semantic differential scale of 1 to 5 between two opposites, as seen in Figure 7. Each participant's survey results are aggregated, and the outcome variable is represented by the resulting floating-point rating. The survey is also available in French, German, and Spanish for extra language support. Translations are obtained from Bartneck, The Godspeed Questionnaire Series (2008).

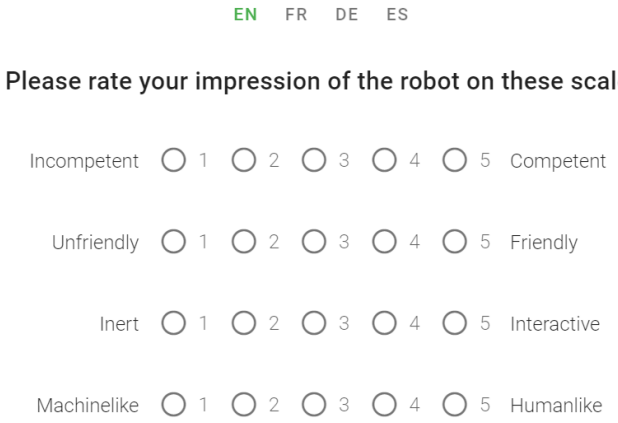

Figure 7: The survey on the web site. It includes 20 items from the Godspeed questionnaires, of which 5 relate to anthropomorphism. The other aspects included are animacy, likeability, and perceived intelligence (5 items each).

Procedure

To begin with, participants declare that they understand and acknowledge the terms of the study (see Figure 8). Their web browser needs to support the required application programming interfaces (see Figure 9 and Figure 10). Since the videos are downloaded over the network, participants need to have a good internet connection. Their device also needs to have a functioning front camera. Before starting, participants authorize access to their camera and run a test to verify that the AFFDEX SDK detects their face (see Figure 11 and Figure 12). Finally, participants provide their age and gender before moving on to the first video.

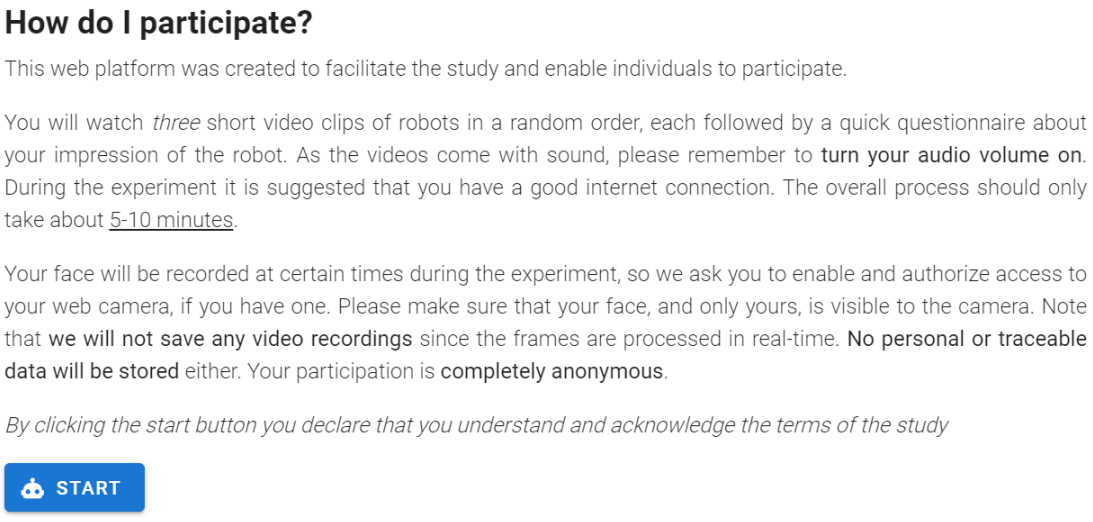

Figure 8: The home page of the web site – introduction, terms, and “start” button.

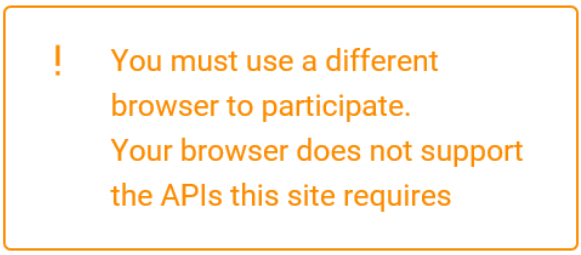

Figure 9: Alert which is displayed on the home page instead of the "Start" button, when the browser doesn't support required APIs, for example, Internet Explorer.

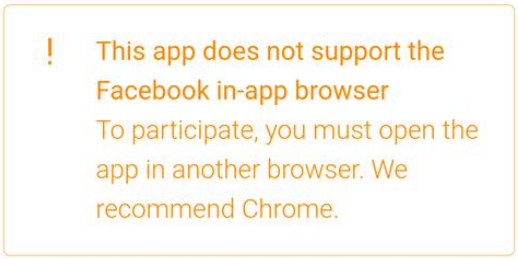

Figure 10: Alert displayed on the home page, instead of the "Start" button, when the user opens the web site in Facebook mobile apps. In-app browsers are only webpage viewers and don't support all the other functionalities that "normal" browsers have built-in.

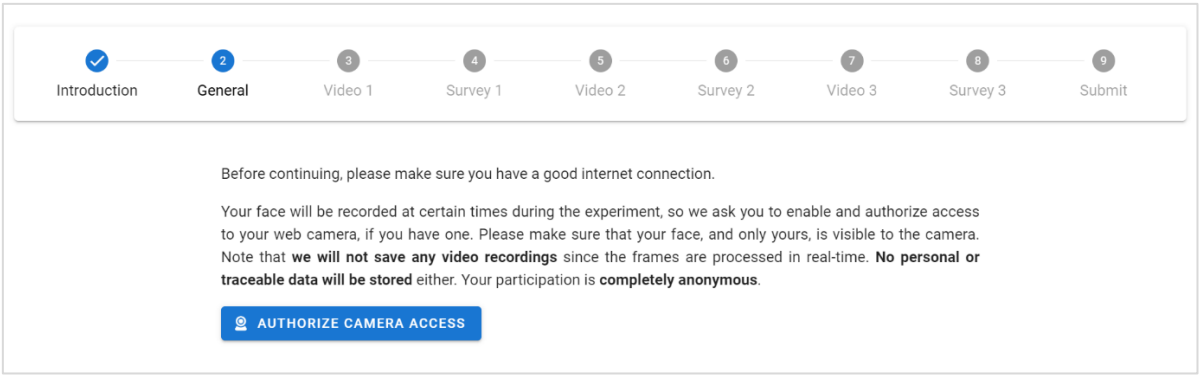

Figure 11: Before moving on, the participant needs to allow the site to control the device's camera.

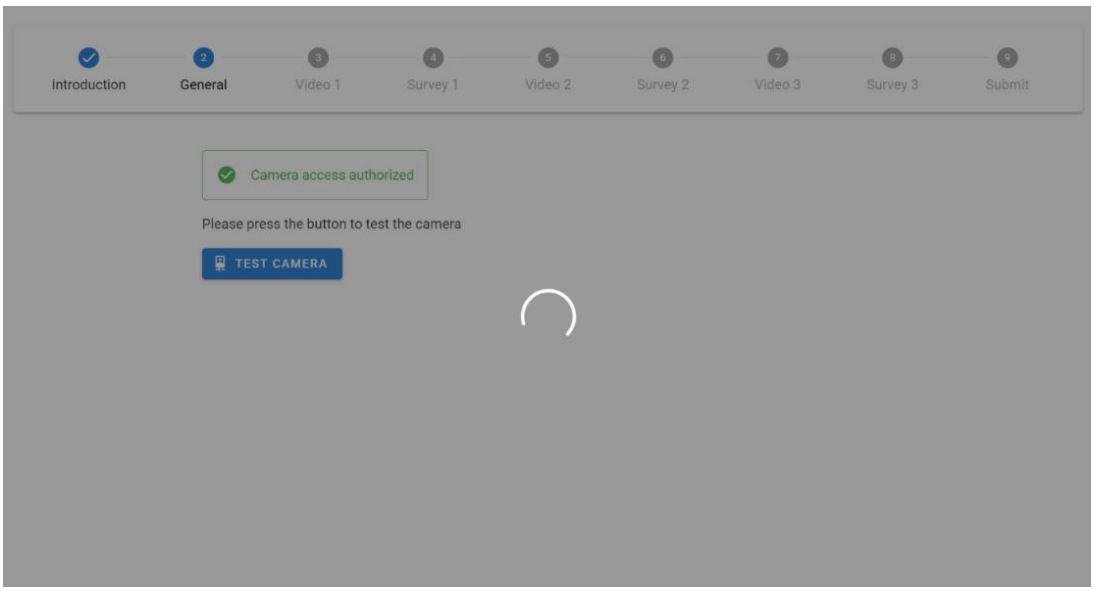

Figure 12: Face detection needs to be tested. This takes approximately 10-30 seconds. When this step is complete, the participant provides their age and gender. Thereafter, they can move on to the first video.

The video clips are played in random order. This is done to avoid response bias that would result from a fixed order based on the level of human likeness, such that participants would answer the questionnaire to meet the expectations of the study. While watching a video, the participant's camera stream is processed at a rate of 30 frames per second by a web worker program running in a background thread. This is done so that the heavy processing does not slow down the user interface (Web Workers API, 2020). The web worker calls the AFFDEX SDK which, in turn, predicts the facial expressions of each participant. If the system is unable to detect a face, a short message is displayed to the participant, thereby reducing the risk of data loss (see Figure 13). The video playback is fully controlled by the system and manual controls are disabled so that the participant cannot pause or stop the video at any moment.

When participants have finished the experiment, they answer the question "Have you ever interacted with a robot in real life?" before submitting the data for permanent storage.

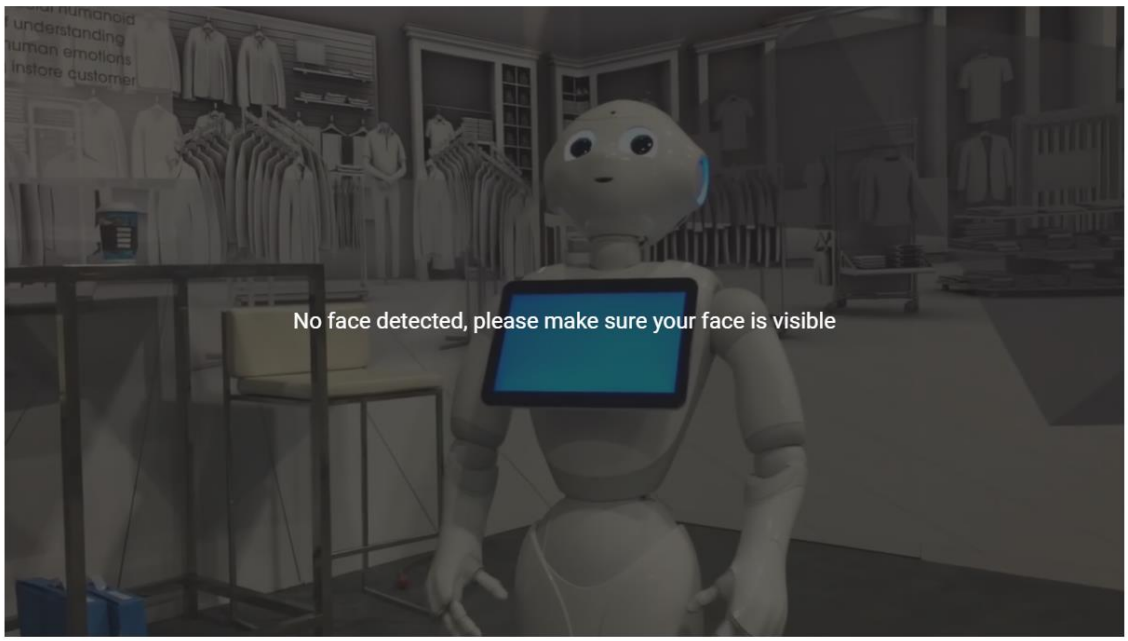

Figure 13: The participant is notified when their face is not detected.

It is difficult to objectively determine the absolute feeling of an individual since the survey answers can be biased by assumptions and prior experiences. Bartneck et al. (2008) also note that when a participant reflects on the experience following some interaction, certain biases can influence their survey response. To partially address this, all survey items are randomized to mask the intention of the survey.

Statistical Analysis

All data processing and model training is implemented in the Python programming language. Pandas, an open-source tool built on top of Python, is used for data processing and exploration, as it is widely used and convenient for this purpose (Pandas, 2020). Scikit-learn, also a popular open-source tool built on top of the same libraries like Pandas, is used for predictive analysis (Scikit-learn, 2020).

While each video clip is playing, facial action scores are collected by the system. The distribution of non-zero facial action scores is sparse for most actions. This is to be expected since sparsity is inherent to natural and spontaneous facial expressions (Senechal, McDuff, & el Kaliouby, 2015). That means that most of the time, people don't show any strong expressions, and possibly even more so when looking at their devices. To handle the sparsity, facial action scores need to be aggregated carefully to produce the predictor variables (features). Taking the average does not make sense for the kind of sparsity we are dealing with here since all aggregated values would be close to zero. Here, we compare two aggregation schemes: 1) taking the max value, and 2) adding up positive slopes, i.e. facial action upswings.

Pearson correlation analysis is performed to check pairwise correlations of features as well as correlations of each feature with the survey response. It is expected that facial actions will be correlated with each other to some degree. We are most interested in the correlation of features with the anthropomorphic response, but for curiosity, we also check correlations with

the overall Godspeed scale and the other measured Godspeed subscales: animacy, likeability, and perceived intelligence. Also, we investigate if there is a noticeable difference in how participants perceive each robot, both in terms of facial actions and survey responses.

Finally, we attempt to train a prediction model using Lasso regression. Lasso is a linear method that incorporates L1-regularization or shrinkage to reduce over-fitting. The features are not expected to all be good indicators for the survey response. Therefore, Lasso is ideal since it produces a sparse model with fewer non-zero coefficients than other linear models (Jaggi & Urbanke, 2019). In other words, feature selection is inherent in Lasso as it enforces sparsity in the model in that some model coefficients will be strictly zero. Retaining only relevant features produces a model that has higher interpretability and potentially lower prediction error (Hastie, Tibshirani, & Friedman, 2017). K-fold cross-validation is applied to further reduce over-fitting and enhance the model selection process.

We report the coefficient of determination or R-squared of the final model. R-squared is an evaluation metric that captures how well the resulting model performs compared to the naïve constant model that always predicts the expected value (mean) of the response variable, regardless of the input features (Scikit-learn, 2020). The best possible score is 1 and the constant model would get a score of 0. The score can be negative, in which case the model performs worse than the constant model.

RESULTS

Over two weeks, 47 people participated in the study, 27 males and 20 females. As summarized in Table 1, around 38% (18/47) of participants have interacted with a robot in real life. Figure 14 reveals that most participants are 17-45 years old and the largest age group is 20-25 with 22 participants.

		robotRealLife		All
		No	Yes	
gender	Female	14	6	20
	Male	15	12	27
All		29	18	47

Table 1: Frequency table showing the gender vs. whether participants have interacted with a robot in real life.

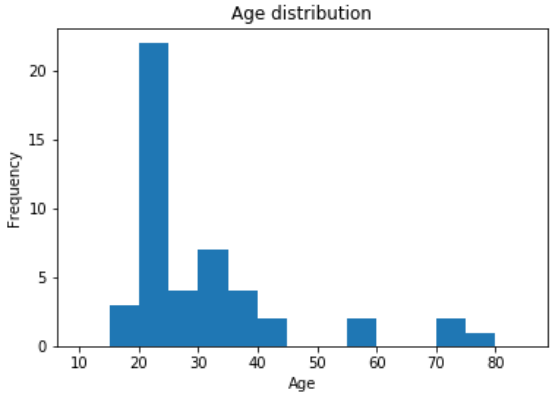

Figure 14: Age distribution of participants

We consider as features 14 facial actions predicted by the AFFDEX classifier. Figure 15 and Figure 16 reveal how they correlate with the survey response. No significant correlations can be found between any of the features and the response, the range being [-0.29, 0.20] for max-aggregated features and [-0.24, 0.13] for slope-aggregated features.

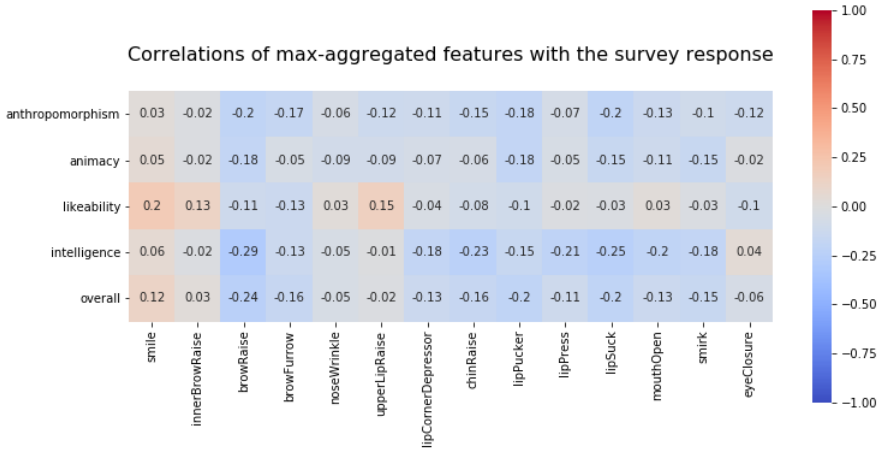

Figure 15: Correlations of max-aggregated features with different levels of the survey response

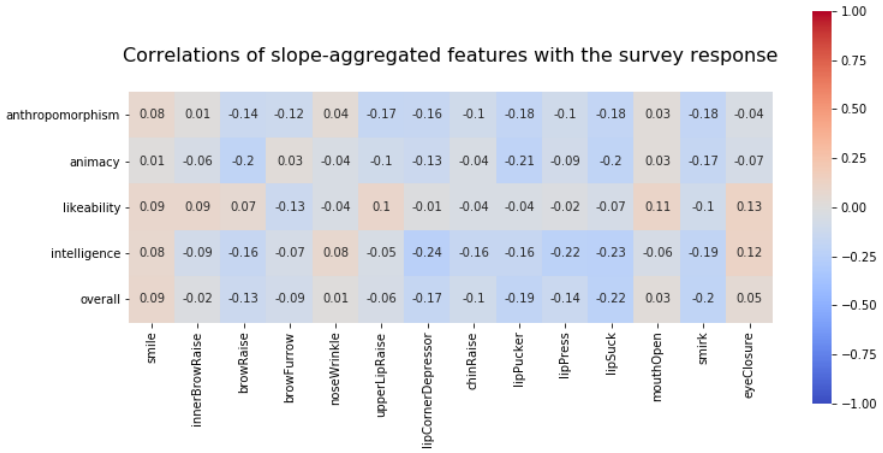

Figure 16: Correlations of slope-aggregated features with different levels of the survey response

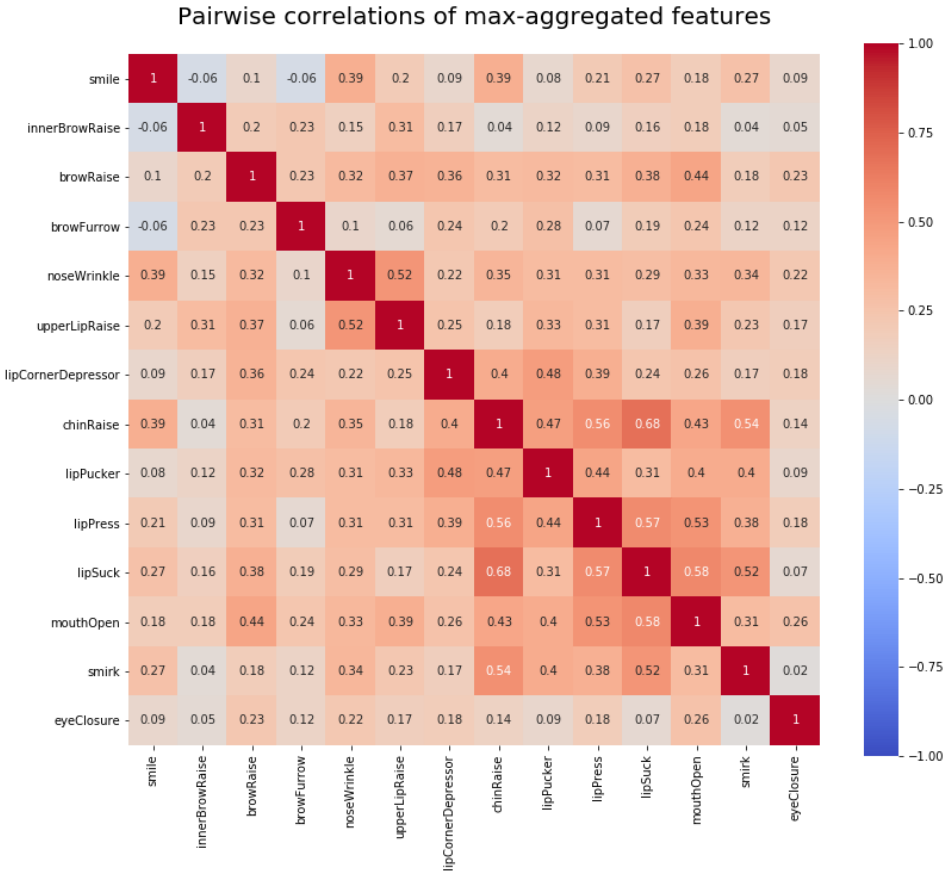

Figure 17: Pairwise correlations of max-aggregated facial action features

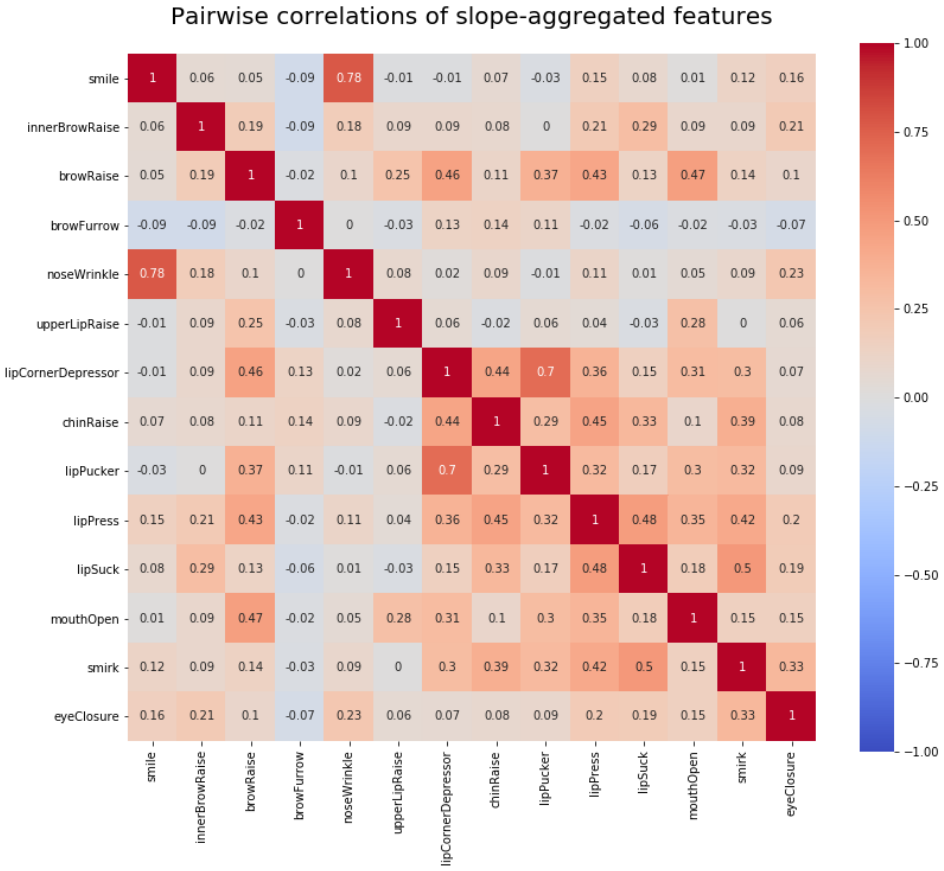

Figure 18: Pairwise correlation matrix of slope-aggregated facial action features

As expected, most of the features are positively correlated with each other to varying degrees (see Figure 17 and Figure 18), especially the max-aggregated features. In layman's terms, facial actions influence one another as they are all part of a single face. Statistically, this implies that the linear relationship between the intensities of two facial actions is positive. A higher correlation implies a stronger positive linear relationship.

For the sample in question, some differences were observed among the robots in terms of levels of facial actions (see Figure 19 and Figure 20). For instance, furrowed brows were less common for Sophia than the other two robots. However, the differences are not statistically significant as can be inferred from the confidence intervals.

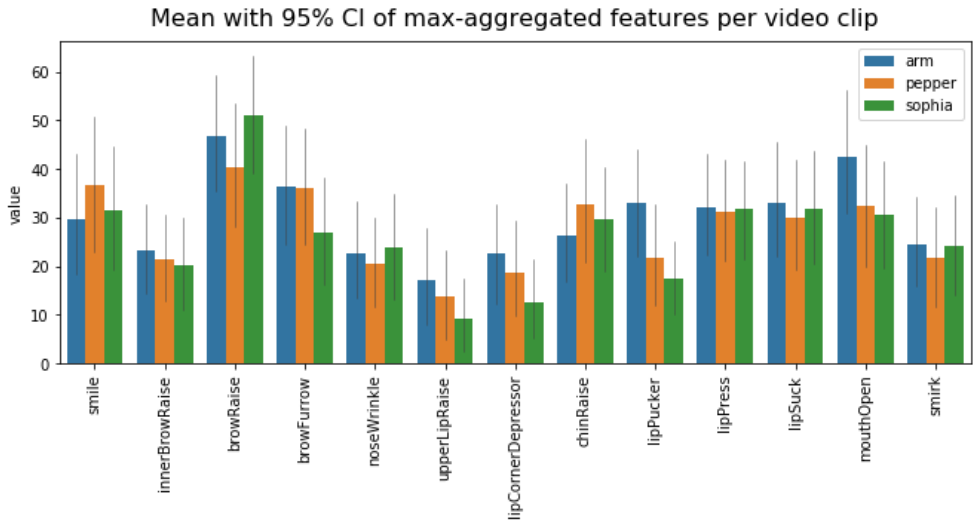

Figure 19: Average of max-aggregated facial actions per video clip

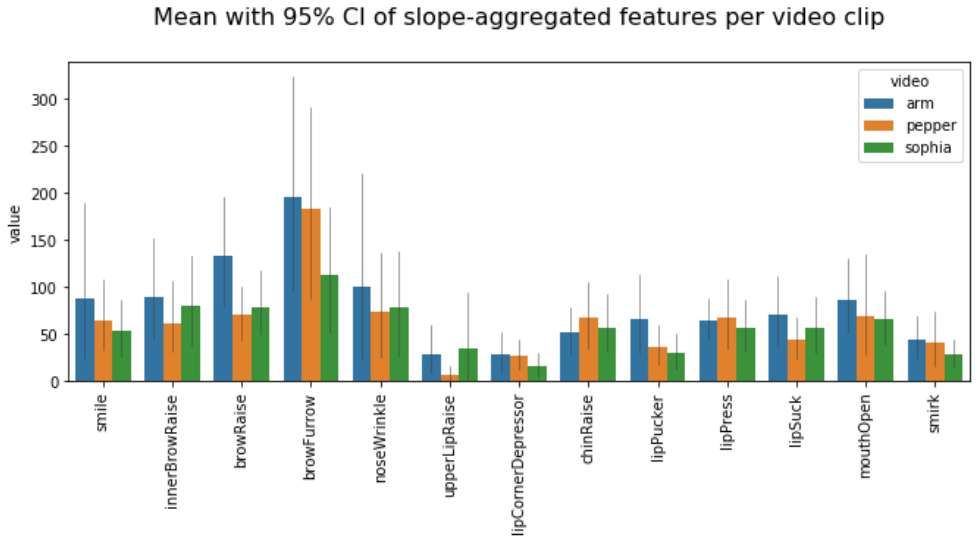

Figure 20: Average of slope-aggregated facial actions per video clip

Possibly more interpretable differences are observed when looking at the average survey response for each video as displayed in Figure 21. Although most of the differences are not statistically significant, the results are interesting. Participants found the robot arm to be the least anthropomorphic and Sophia to be the most anthropomorphic. Curiously, the converse can be said about perceived intelligence. There is a significant difference between how participants perceived the animacy or liveliness of the robot arm compared to Pepper and Sophia. Participants also found Sophia to be the least likable.

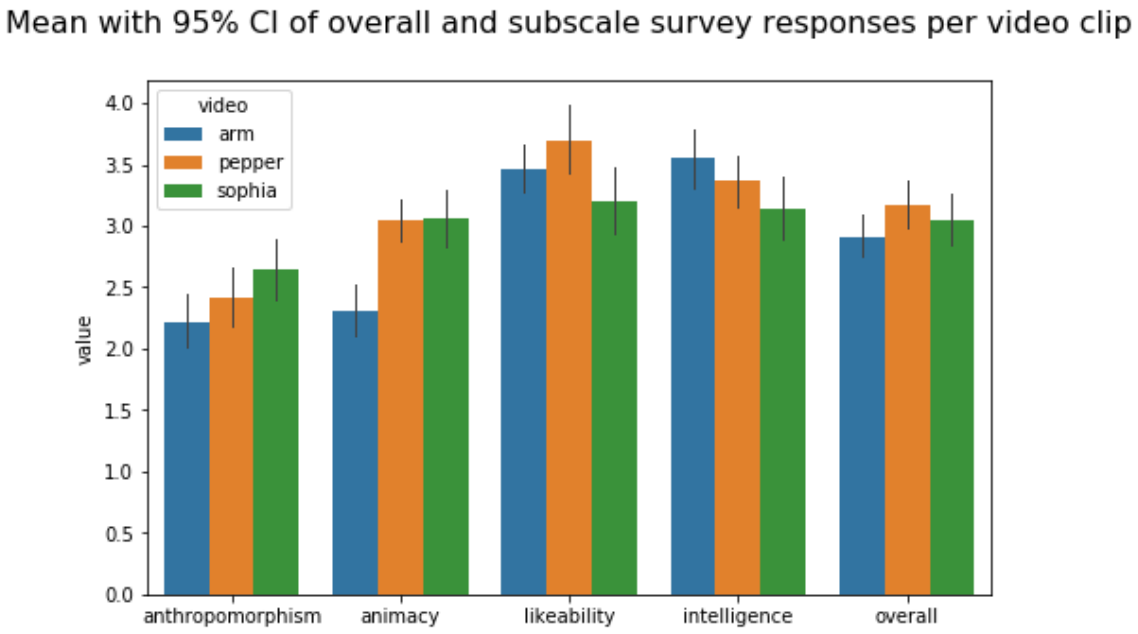

Figure 21: Average of survey responses, overall and subscales, per video clip

To understand better how well video clips can be inferred from the facial actions and survey responses, we predict the video clips using a Gaussian Naïve Bayes model and computed the average accuracy of 100 different train-test splits. For max-aggregated and slope-aggregated facial actions, the average accuracy is 26.6% and 31.1% respectively. Using the response to each question in the survey as features, we report an average of 71.7%.

The data samples were split into training and test sets to fit models to predict survey responses from the facial expressions. Training samples were fit to linear models with Scikit-learn's LassoCV (Lasso with cross-validation). Both max-aggregated and slope-aggregated

features were used and the R-squared score of each model was computed after applying the best model to the held-out test samples. The results can be gleaned from Table 3 and Table 2. Most of the R-squared values are slightly less than zero, which implies that the models perform slightly worse than the naïve model that always outputs the mean of the response. This is explained by the fact that the non-zero coefficients of the final models are in most cases 0 and at most 1 out of 14. In other words, the method reduces most feature weights to zero, implying that they are useless for predicting the response. This is not surprising since we observed no significant correlations between the features and the response (Figure 15 and Figure 16). The model is constant in the case when all the coefficients are zero.

	R-squared
anthropomorphism	-0.232290
animacy	-0.088932
likeability	-0.007692
intelligence	0.027311
overall	-0.057900

Table 3: R-squared of the prediction of Lasso models for *max*-aggregated features

	R-squared
anthropomorphism	-0.232290
animacy	-0.085901
likeability	-0.007692
intelligence	-0.023546
overall	-0.057900

Table 2: R-squared of the prediction of Lasso models for *slope*-aggregated features

The results of this study, including tables and figures, can be reproduced by following the instructions on the GitHub repository.

DISCUSSION

The video clips were intentionally made to be few and short in length to keep the experiment short. Of course, three video clips are not enough to represent the whole range of robots that are out there. The more types of robots we have, the more robust the model will be and better at predicting responses to robots the model hasn't encountered before. Therefore, an improvement could be made by having more video clips and only showing a subset of those to each participant.

We also need to consider the number of samples. As discussed before, facial actions are sparse and quite unpredictable, so we need a greater number of faces to discover general trends in the anthropomorphic response to robots. We can't say for sure if there are undiscovered trends unless we gather more data.

We reported an average accuracy of 71.7% when predicting the video from the responses to individual survey items as opposed to a much lower accuracy when predicting based on facial actions. This was expected since the survey is a more objective measure of people's feelings about robots. The result is also supported by the fact that the selected video clips show robots that are different in terms of human likeness and sophistication, such that people distinguish between them.

When looking at Table 3 and Table 2 it might appear as if the method used to aggregate the facial action scores does not matter in determining the accuracy of the predictive model. However, that's not entirely true because, as explained, the coefficients of the resulting models are mostly zero. This means that the features are simply discarded as "useless" in the final model. So, there is no way to tell which aggregation method is better by using the current samples. However, with more data patterns might emerge and reveal unseen relationships

between facial actions and anthropomorphic attitudes towards robots. Also, more sophisticated aggregation methods could be applied.

Furthermore, it may be argued that aggregating a long time series to produce a single number to represent each facial action and, in turn, using these numbers to predict the person's overall impression of the robot is far-fetched, especially since the end-goal is to have a model that can predict the person's impression instantly in real-time. People's feelings about something are not constant, they rather develop over time and such a model doesn't capture that development. This is related to the "first impression bias," a well-known psychological phenomenon, studied for instance by Asch (1946) and Lim, Benbasat, & Ward (2000). The participant will not be fully cognizant of the robot's capabilities and limitations after only a short period of observation or interaction. We need a model that takes this fact into account and doesn't jump to conclusions too early.

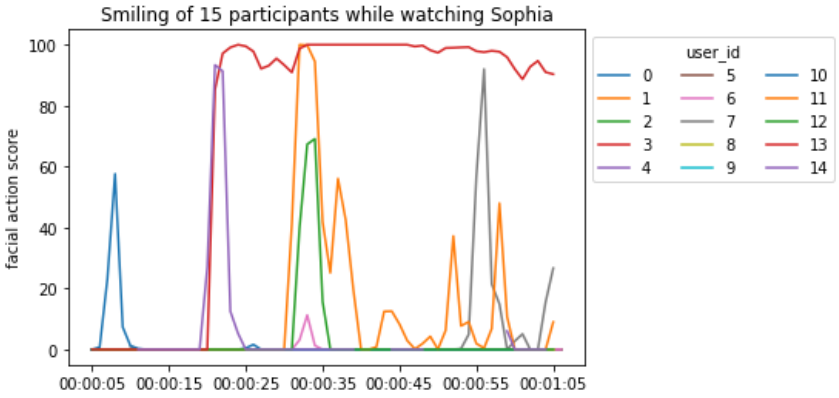

Figure 22: Smile facial action time series of 15 participants during the video of Sophia the Robot

To achieve this, the method could be improved by cutting each video clip into smaller independent clips and surveying the anthropomorphic response after each one. Additionally, a real interaction could be mimicked by using pre-recorded videos of robots asking questions and responding to the participant. This could be implemented by a decision tree where relationships between questions and responses are hard-coded and participants are only given a limited number of options to respond to the robot. Going one step further, participants could interact with a robot through a conversational agent interface (a.k.a. chatbot) where the robot's image is displayed, and the person's impression is measured regularly during the conversation.

CONCLUSION

To conclude, using the data collected in the online experiment, we found that facial expressions are not indicative of the anthropomorphic response towards robots. However, survey responses were found to be good predictors for the video clip, i.e. the type of robot. Aggregated facial action scores were used as features to represent the facial response of each participant towards a given robot. The features were not significantly correlated with the survey response and, as a result, the resulting predictive models perform worse than the naïve constant model. However, this is most likely due to the small sample size. Since facial expressions are sparse, more data is needed to enable the discovery of potential relationships.

REFERENCES

1. Asch, S. E. (1946). Forming impressions of personality. *J. Abnormal Soc. Psych, 41(3)*, 1230–1240. Retrieved from https://doi.org/10.1037/h0055756

2. AUBO Robotics. (2017, May 16). *Chinese Tea Art - AUBO i5 Dual-Arm Robot.* Retrieved from YouTube: https://www.youtube.com/watch?v=3y9N1l7ofYY

3. Bartneck, C. (2008, March 11). *The Godspeed Questionnaire Series.* Retrieved from Christoph Bartneck, Ph.D.: http://www.bartneck.de/2008/03/11/the-godspeed-questionnaire-series/

4. Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2008). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, (pp. 71-81). Retrieved from https://doi.org/10.1007/s12369-008-0001-3

5. Caridakis, G., Castellano, G., Kessous, L., Raouzaiou, A., Malatesta, L., Asteriadis, S., & Karpouzis, K. (2007). Multimodal emotion recognition from expressive faces, body gestures and speech. In P. A. Boukis C. (Ed.), *IFIP International Conference on Artificial Intelligence Applications and Innovations. AIAI 2007: Artificial Intelligence and Innovations 2007: from Theory to Applications*, (pp. 375-388). Retrieved from https://doi.org/10.1007/978-0-387-74161-1_41

6. Dautenhahn, K. (2014). Human-Robot Interaction (section 38). In *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* Retrieved from https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-robot-interaction

7. Eyssel, F., Kuchenbrandt, D., & Bobinger, S. (2011). Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. *HRI 2011 - Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, (pp. 61–67). Retrieved from https://doi.org/10.1145/1957656.1957673

8. funtime 247. (2017, November 10). *Artificial intelligent robot Sophia Journalists interviewing.* Retrieved from YouTube: https://www.youtube.com/watch?v=LQnJqjW8bGY

9. Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications, 30(4)*, 1334-1345. Retrieved from https://dx.doi.org/10.1016/j.jnca.2006.09.007

10. Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. Springer.

11. Jaggi, M., & Urbanke, R. (2019, October 3). Regularization: Ridge Regression and Lasso. *Machine Learning Course - CS-433*. Retrieved from https://github.com/epfml/ML_course/raw/master/lectures/03/lecture03d_ridge.pdf

12. Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. *Extended abstracts of the 2002 Conference on Human Factors in Computing Systems*, (pp. 576-577). Minneapolis, Minnesota, USA. Retrieved from https://doi.org/10.1145/506443.506491

13. Lim, K., Benbasat, I., & Ward, L. (2000). The Role of Multimedia in Changing First Impression Bias. *Information Systems Research, 11*, 115-136. Retrieved from https://doi.org/10.1287/isre.11.2.115.11776

14. Loftsson, V., & Griesser, P. (2020). Retrieved from Anthropomorphic Expressions: https://www.anthropomorphic-expressions.live/

15. MacDorman, K. (2006). Subjective Ratings of Robot Video Clips for Human Likeness, Familiarity, and Eeriness: An Exploration of the Uncanny Valley. *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science.* Vancouver.

16. McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & Kaliouby, R. (2016). AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, (pp. 3723-3726). Retrieved from https://doi.org/10.1145/2851581.2890247

17. MobileSyrup. (2016, March 2). *Pepper the Robot in action - MobileSyrup.com.* Retrieved from YouTube: https://www.youtube.com/watch?v=PqtyXvSva4w

18. Pandas. (2020). Retrieved from Pandas: https://pandas.pydata.org/

19. Powers, A., & Sara, K. (2006). The advisor robot: tracing people's mental model from a robot's physical attributes. *1st ACM SIGCHI/SIGART conference on Human-robot interaction.* Salt Lake City, Utah, USA. Retrieved from https://doi.org/10.1145/1121241.1121280

20. Scikit-learn. (2020). Retrieved from Scikit-learn: https://scikit-learn.org/stable/

21. Scikit-learn. (2020, June 13). *3.2.4.1.3. sklearn.linear_model.LassoCV.* Retrieved from Scikit-learn: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoCV.html#sklearn.linear_model.LassoCV.score

22. Senechal, T., McDuff, D., & el Kaliouby, R. (2015). Facial Action Unit Detection Using Active Learning and an Efficient Non-linear Kernel Approximation. *ICCVW '15:*

Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop, (pp. 10–18). Santiago. Retrieved from https://dl.acm.org/doi/10.1109/ICCVW.2015.11

23. Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7(3)*, 487–494. Retrieved from https://doi.org/10.1037/1528-3542.7.3.487

24. *Web Workers API.* (2020, June 9). Retrieved from MDN web docs: https://developer.mozilla.org/en-US/docs/Web/API/Web_Workers_API

Figures

Tables