2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)
Beijing, China, November 6-9, 2018

Can a Humanoid Robot Spot a Liar?*

A.M. Aroyo, J. Gonzalez-Billandon, A. Tonelli, A. Sciutti, M. Gori, G. Sandini and F. Rea

***Abstract* — Lie detection is a necessary skill for a variety of social professions, including teachers, reporters, therapists, and law enforcement officers. Autonomous system and robots should acquire such skill to support professionals in numerous working contexts. Inspired by literature on human-human interaction, this work investigates whether the behavioral cues associated to lying – including eye movements and response temporal features – are apparent also during human-humanoid interaction and can be leveraged by the robot to detect deception. The results highlight strong similarities in the lying behavior toward humans and the robot. Further, the study proposes an implementation of a machine learning algorithm that can detect lies with an accuracy of 75%, when trained with a dataset collected during human-human and human robot interaction. Consequently, this work proposes a technological solution for humanoid interviewers that can be trained with knowledge about lie detection and reuse it to counteract deception.**

I. Introduction

Generally speaking deception is the act of hiding the truth using a false statement with the intention to make someone else believe it. The intentions behind deception can be several and can be gathered into two main groups: cooperative or explorative intentions. Cooperative deception could be defined as a lie with the goal to protect someone (feelings, interests) that can bring to an enhancement of social bounds. Conversely, exploitative deceptions are used by the manipulators to exploit vulnerabilities for their own benefit.

In modern contexts, deception has a relevant impact on social activities particularly those that require tutoring (*e.g.*: in educational programs and healthcare). The ability to detect deception is a necessary skill for a broad range of professions, including teachers, reporters, therapists, and law enforcement officers. Such professionals are usually trained to detect deception in order to tailor their professional activity to the specific individual's predisposition to lie. By detecting deceit, experts increase their emotional distance between themselves and the interviewed, while recognize distrust.

Unfortunately, artificial intelligent systems are far from valuing to operate similarly, by identifying deceits in order to prepare attuned and opportune intervention strategies, as competent professionals do. Autonomous systems can rely on different cues that have been proved to be altered by the cognitive processes at the basis of deception. Traditional automated methods used for lying detection (*e.g.,* polygraph, heartbeat sensor, blood pressure monitor, sweat and respiratory rate measurement devices) are invasive and require an experienced human interviewer. Although, recently, other cues have attracted considerable attention as relevant lying indicators because of their immediate portability on autonomous systems and reduce invasiveness.

It has been showed that lying can require more cognitive load compared to telling the truth [1], [2]. For example, liars need to build a plausible story and monitor its coherence [3]–[6]. Moreover, liars are more inclined to monitor and control their behavior and as well as the behavior of the "interviewer". Recent evidences in the literature [7]–[15] propose a direct link between lie preparation and oculomotor patterns such as blinking, fixations, saccades and pupillary response.

In fact, eye blinking and pupil dilatation are usually associated to cognitive load processing [16]. It was reported that the time interval between the onset of a stimulus and the blinking onset is delayed by cognitive processes and motor responses [17], [18]. Leal and Vrij [7] tested that hypothesis recording the frequency of blinking while lying or telling the truth. When saying a lie, the blinking pattern exhibited by participants was strikingly different from the one obtained by telling the truth. In particular, liars showed a decrease in eye blinks while uttering the lie, followed immediately by a substantial increase in blinking frequency.

The pupillary response seems a highly sensitive instrument for tracking fluctuating levels of cognitive load. Beatty and Lucero-Wagoner [19] identified three useful task-evoked pupillary responses (TEPRs): mean pupil dilation, peak dilation, and latency to the peak. Another example of the importance of the pupillary response has been provided by Dionisio *et al.* [20], they asked students to reply to questions, sometimes saying the truth and other times telling a lie. The task-evoked pupil dilatation was significantly greater when participants were confabulating responses compared to when they had to say the truth about an episodic memory. These results suggest that the increased pupil size could be associated with a deceptive recall.

In another experiment, Walczik *et al.* [21] decided to test whether elaborating deceptive answers can be correlated to the time to respond. They discovered that the decision to lie adds time in the response, especially in open-ended questions (*i.e.,* questions that elicit more than two possible answers).

Notably, robots are also starting to be used in the context of professional activities requiring deception detection skills, such as in security, education, or healthcare. However, differently, from non-physically present autonomous systems,

*This work was supported by RoboCom++ FLAG-ERA Joint Transnational Call (JTC) 2016 Project.

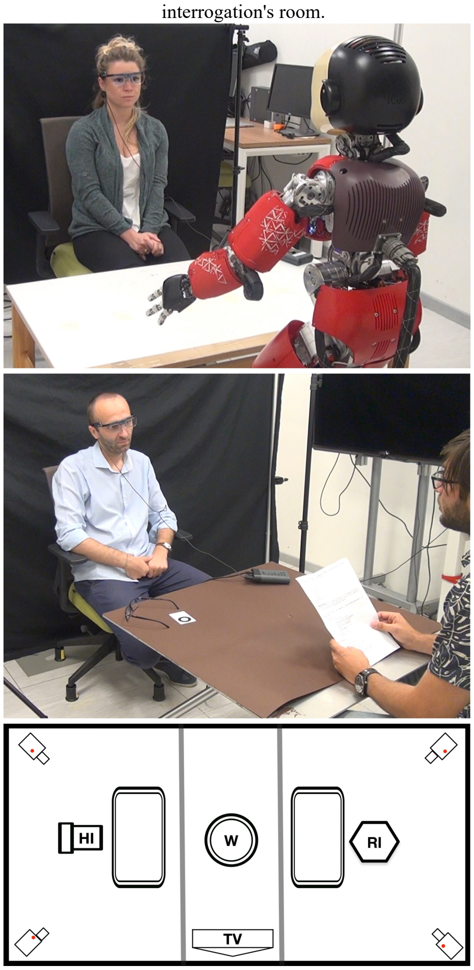
A. M. Aroyo and J. Gonzalez-Billandon are with DIBRIS, Università di Genova, Opera Pia 13, 16145, Italy and with RBCS, Istituto Italiano di Tecnologia, Enrico Melen 83, Bldg B, 16152 Genova, Italy alexander.aroyo@iit.it and jonas.gonzalez@iit.it.

A. Tonelli and M. Gori are with U-VIP, Istituto Italiano di Tecnologia, Enrico Melen 83, Bldg B, 16152 Genova, Italy alexander.aroyo@iit.it and alessia.tonelli@iit.it and monica.gori@iit.it.

A. Sciutti; G. Sandini and F. Rea are with RBCS, Istituto Italiano di Tecnologia, Enrico Melen 83, Bldg. B, 16152 Genova, Italy. alessandra.sciutti@iit.it, giulio.sandini@iit.it and francesco.rea@iit.it.

robots can take advantage of their embodiment [22]. Recent research proves that the physical presence of others has an effect on increasing the cognitive load during the deception [23] and inducing cognitive load has been suggested as a valuable strategy to facilitate lie detection through the assessment of response time, answer consistency, eye movements and pupil dilatation [10]. Further, the humanoid appearance constitutes an additional element that might influence the level of cognitive load. Within fact, the humanoid shape might trigger a process of anthropomorphisation leading to ascribe to the robot similar capabilities and psychological features as those of a human interviewer [24], [25].

Figure 1. Top - the robot investigator (RI) debriefing a witness (W); middle - human investigator (HI) debriefing a witness (W); bottom - interrogation's room.



This work investigates the possibility to detect deception in a human-humanoid interaction, by monitoring behavioral cues proven to be significantly affected by telling lies in presence of a human interviewer. This study considers collaborative deception, asking participants to lie to protect someone else. The purpose of the study is to be able to recognize lies defined as the attempt to make another agent believe as true propositions which are actually false. To this aim, it is first assessed whether an interview performed by a humanoid robot elicits the same responses as one performed by a human agent. Moreover, it is assessed a possible implementation of a machine learning solution that could be adopted by an interactive robotic platform to detect deception in natural interactions.

II. Methodology

A. Participants and Experimental Design

15 participants were recruited from the institute, 60% females with an average age of 36.47 years (SD=12.68) with a broad educational background. All of them participated for free to the experiment and signed an informed consent form approved by the local ethical committee, in which, it was stated that camera and microphone can record their performance, and agreeing on the use of their data for scientific purposes.

Participants were equally distributed among a 2x2x2 conditions to avoid any ordering effects (agent: human or robot investigator; witness: truth-teller or liar; and two different videos). The agent order was kept constant within the same condition. Before starting the experiment, participants were asked to avoid drinks with caffeine and stimulating substances, to preserve normal physiological alteration.

B. Setup

For the purpose of the experiment, the experiment room was prepared as an interrogation room (Figure 1, bottom). The room was divided into three zones with black curtains, with the witness (W) seated in the center on a rotating chair. This setup allowed to quickly switch from robot (RI) to human interviewer (HI) and also to ensure complete isolation during the interrogation. The cameras, placed in the corners, were used to record the participant during the whole interrogation. 4K and HD cameras were used to record the participants when they were interviewed, together with the chest sensor and an ambient microphone. Participants wore Tobii Eyetracker glasses. These glasses, with a frequency of 100Hz, were used to record pupil dilatation and eyes movements. To ensure same setup for all the participants during different times of the day, the windows blinders were closed, and the room was lit with artificial light. This also guaranteed the pupil dilatation to be similar across the participant population.

C. Procedure and Materials

The experiment is inspired in the work of Walczik *et. al,* [10] and it is divided into three phases: (i) general questions, (ii) first question session, (ii) second question session. Before starting the experiment, participants were asked to fill in a questionnaire (fully described below) to identify particular psychological features that could influence the results.

Participants were welcomed by an experimenter, who explained the general purpose of the experiment: "They have been witnesses of two crimes, and they have to help the investigators to find out the responsible".

Once in the room, the experimenter asked the participants to wear the Tobii Eyetracker glasses and Polar H10 heart-rate chest sensor.

The experimenter asked the participant to sit in the middle of the room and calibrated the Tobii glasses. After, the participant was instructed to answer truthfully and quickly to 20 general questions (*e.g.,* "Can an oven be hot?", "What is the first name of Berlusconi?"), the first and the second question session were initiated. Respectively the first 10 questions were asked by the robot investigator and the remaining 10 by the human investigator (Figure 1 - top, middle). The order of the block of human and the robot questions was alternated within the participants. In the room a black curtain separated the participant and the investigator from the inactive investigator. The experimenter always left the interrogation room before the investigator started the questions.

After completing the first phase of general questions, the experimenter entered the interrogation room and gave the participant an instruction sheet. It was written that s/he was the witness of a crime, and s/he should pick "randomly" a role from a box (the randomization was just an illusory effect for the participant since the role was defined a priori). Inspired by [10], the role could either be: *truth tellers* - a witness who wants the criminals brought to justice, thus, to reply to all the questions truthfully; or *protectors* - a witness who realized that the criminal is a familiar of theirs, and should lie to all the questions in order to protect the familiar. Participants were asked to be coherent and reply deceptively to all the questions.

One video, shown only once in the presence of the experimenter on the TV screen (Figure 1 - bottom), was 59s in length and featured three mid-aged white males in an empty clothes store. The perpetrators communicated to each other using signs, one of them opened a paper bag and the other put inside different types of clothing. After, the three of them left the shop serenely. The other video, of 101s in length presented a white male teenager dressed in sportive clothing with a hat and skateboard. He was loitering in an electronic shop while the cashier was attending another client. At some point, the teenager picked a game CD, went behind the stands and tried to put the game in his pants.

After the participant watched the video, the experimenter put and calibrated the Tobii glasses again, and left the interrogation room to the investigators. Either the robot or human, asked in turn 10 questions each, in the two different locations of the room (Figure 1 - bottom). The investigators made two types of questions: short type - yes/no questions, and open-ended questions. A sample of a short question was "Was the criminal dressed in a formal way?"; while the open-ended was "How did the criminal hide the loot?".

These questions differ in syntactic constraints that put on permissible responses [26]. An example of a short question is "Did it happen in a crowded place?"; while an open-ended one is "What was the criminal act?".

After both interrogations, during the last phase, the experimenter entered in the room again and made the witness pick "randomly" a role (the condition was forced to be the opposite of the previous one).

When both interrogators finished, the experimenter entered in the room to remove the glasses, and to ask the participant to compile a final questionnaire before finishing the experiment. At the end of the experiment, the experimenter removed the heart-rate chest sensor and accompanied the participant to leave the room.

D. Measurements

The measures are separated into the following categories:

Questionnaires: (i) demographic statistics such as gender, age, nationality and education; (ii) the 60 item Big Five personality traits [27]; (iii) (vi) the Negative Attitude towards Robots Scale (NARS) [28]; (iv) Brief Histrionic Personality Scale (BHPS) [29]; (v) Dark Triad of Personality Short [30].

Behavioral measures: (i) time to respond (from the moment the investigator finished the question till the witness started replying); (ii) eloquence time (time the witness spend replying to the question); (iii) number of saccades; (iv) number of fixations; (vi) number of blinks; (vii) left and right pupil dilatations - max, min and average.

III. RESULTS

One participant had to be removed as he did not understand the instructions of the role adaptation - he ended up adopting the same role for both videos. Regarding the data analysis, the outliers were filtered inspired by [10], [31].

The results of the psychological profile can be found in Table 1. The personality profile of the participants is well distributed in the sample without having high values for the Histrionic and Dark Triad scales.

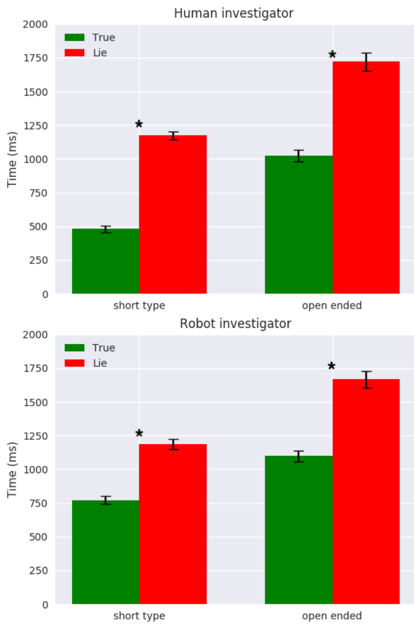
TABLE I. NARS: S1, S2, S3 - HIGHER MORE NEGATIVE; BIG 5: EXTRAVERSION, AGREEABLENESS, CONSCIENTIOUSNESS, NEUROTICISM, OPENNESS TO EXPERIENCE - THE HIGHER THE STRONGER; HISTRIONIC - HIGHER MORE NEGATIVE; DARK TRIAD: MACHIAVELLIANISM, NARCISSISM, PSYCHOPATHY - THE HIGHER THE STRONGER. THE NUMBERS REPRESENT ARE THE NUMBER OF SUBJECTS EXHIBITING THE CORRESPONDING SCORE %

Score %	Participants' psychological profile			
	NARS	*Big 5*	*Histrionic*	*Dark Triad*
0-20%	{7,5,3}	{1,0,0,1,0}	1	{1,1,7}
20-40%	{4,5,3}	{4,1,1,7,0}	6	{8,5,5}
40-60%	{2,3,5}	{3,3,6,4,6}	6	{4,8,2}
60-80%	{1,1,2}	{5,7,6,2,6}	1	{1,0,0}
80-100%	{0,0,1}	{1,3,1,0,2}	0	{0,0,0}

Two two-way repeated measure ANOVAs were performed on all the features of the vector for short and open-ended question types respectively. The factors were Veracity (truth/lie) and Interviewer (human/humanoid). There is a statistically significant difference in the response time when participants have to say the truth with respect to saying a lie: short type $F(1,13)=34.66$, $p<0.001$; open-ended $F(1,13)=16.1$, $p<0.001$ (Figure 2).

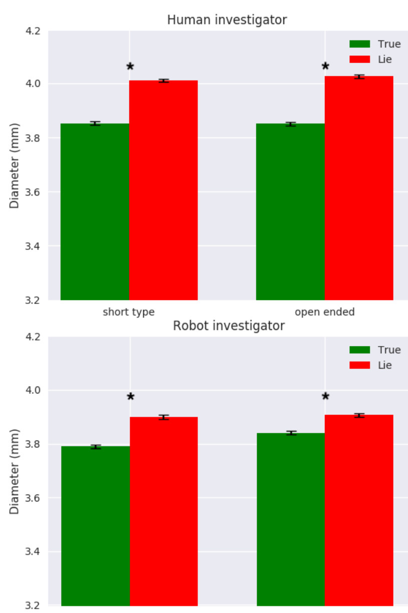
There are also significant differences in the average pupil dilatation for both eyes while telling the truth rather than a lie. Right pupil: $F(1,13)=10.03$, $p=0.007$ for open-ended questions and $F(1,13)=14.27$, $p=0.002$ for short type. Left pupil: $F(1,13)=7.44$, $p=0.017$ for open-ended questions; and $F(1,13)=12.58$, $p=0.003$ for short type (Figure 3). There is not statistical difference between the dilatation of the right with respect to the left pupil.

Figure 2. Time to respond for short type and open-ended questions when saying the lie or truth to the human or robot investigator. Statistically significant items marked by *.



There were no significant differences between the time participants spent talking with the robot or human investigator (Figure 4).

Figure 3. Average pupil dilatation - left and right pupil. Statistically significant items marked by *.



A way to understand this lack of effect is to analyze the self-reported post-questionnaire where participants had to rate the difficulty to lie to the human and robot investigator. Participants rated 4.29 (SD=0.82) / 10 (very difficult) to lie to the robot, while 5.07 (SD=0.71) to the human counterpart. These items are statistically different with a paired *t*-test t(13)=2.16, p=0.04.

In order to find a connection between the psychological profile of the participants and the behavioral cues, a regression analysis was run on the difference between lie and truth per each participant. Eloquence correlates with openness to experience with F(1,12)=6.13, p=0.19. Another behavioral trait that correlates with the time to respond is neuroticism F(1,12)=8.37, p=0.013.

TABLE II. AVERAGE VALUES FOR ALL THE DIFFERENT FEATURES USED. HI: HUMAN INTERVIEWER, RI: ROBOT INTERVIEWER

Features (short, open end)	**Experiment's data (SD)**			
	HI-Truth	***HI-Lie***	***RI-Truth***	***RI-Lie***
number of saccades/s	5.44(0.18), 8.08(0.32)	9.16(0.22), 11.11(0.41)	7.27(0.36), 11.21(0.53)	7.77(0.23) 9.89(0.29)
number of fixations/s	4.91(0.14), 7.22(0.26)	7.20(0.16), 7.46(0.26)	5.88(0.29), 9.05(0.41)	6.53(0.19), 8.01(0.22)
number of blinks/s	1.26(0.06), 1.61(0.08)	1.73(0.06), 2.46(0.14)	1.71(0.09), 2.33(0.14)	1.74(0.07), 2.16(0.09)
pupil left max (mm)	4.37(0.02), 4.48(0.04)	4.74(0.03), 4.84(0.03)	4.70(0.03), 4.73(0.04)	4.77(0.04), 4.66(0.04)
pupil right max (mm)	4.38(0.02), 4.39(0.02)	4.52(0.02), 4.54(0.02)	4.42(0.03), 4.49(0.03)	4.55(0.02), 4.49(0.02)
pupil left min (mm)	3.35(0.01), 3.31(0.02)	3.32(0.02), 3.38(0.02)	3.26(0.02), 3.30(0.02)	3.33(0.01), 3.37(0.01)
pupil right max (mm)	3.44(0.01), 3.36(0.01)	3.43(0.01), 3.42(0.02)	3.34(0.01), 3.33(0.02)	3.41(0.01), 3.35(0.02)

The rest of the results of the studied features are represented in the Table II. No other differences were significant.

Figure 4. Eloquence for short type and open-ended questions when saying a lie or the truth to the human or robot investigator.

IV. MACHINE LEARNING SYSTEM

The problem addressed in this paper is to test the possibility to train a model from behavioral responses associated with deception and to transfer it into a robotic autonomous system in order to identify true or false answers. This is a binary classification problem, defined by an input vector X; and Y ∈ [0: True; 1: Lie] as desired output vector. The dataset D{X, Y} used to train the model was extracted from the data gathered in the experiment. It was split into the sub-datasets D1 and D2 in order to address two different levels of lie: (i) detect future lies on known participants p1; (ii) detect lies on unprecedented participants p2.

D1 is a set of participants' answers without any participant identification; D2 is a set of participants with their answers, and each answer is associated to the corresponding participant. The D1 dataset is used to demonstrate the possibility to spot future lies from already known participants, but not on unseen participants. The D2 dataset is instead used to try to obtain a general lie detector using a subset of participants as test set.

The literature on deception detection and evidences from the post-analysis provide a starting point to select the features that can be used to create the input vector X. In previous studies [7], [12], [32], eyes features have been shown to be significant to discriminate between lies and true statement, including pupil dilation, number of saccades, fixations and blinks. Speech temporal features as time to respond, and eloquence (*i.e.* the time that the person spends to reply) seem to be also significant for lie detection [10].

The literature review, together with the previous analysis, motivates the selection of the following 11x1 input vector X: [eloquence (milliseconds), time to respond (real in milliseconds), average pupil diameter left (real in millimeters), average pupil diameter right (real in millimeters), max pupil diameter left (real in millimeters), max pupil diameter right (real in millimeters), min pupil diameter left (real in millimeters), min pupil diameter right (real in millimeters), number of saccades (whole number), number of fixations (whole number), number of blinks (whole number)].

To detect future lies of a known person, a decision tree classifier was trained on D1. Decision trees are useful for identifying important features in the data making them transparent and easier to understand. The learned tree can be directly translated into a set of rules to produce an expert system that can be easily ported, for example, into the iCub robotic platform. The second algorithm used was a multi-layer perceptron (MLP) with one hidden layer. A binary cross entropy loss function with Adam optimization was used to train the network. The MLP was trained on D2 to demonstrate the possibility to spot lies answer on unseen person. Furthermore, previous study has demonstrated the possibility to use decision tree to detect lies [33].

Table III shows the different accuracies achieved by the best model of the two algorithms. Due to the small sample size a cross validation has been run for both algorithms to assess the reliability of the results.

Accuracies are shown with the different data augmentation. Both of the two algorithms were trained with a grid-search optimization to find the best hyper-parameters.

Prediction accuracy of a decision tree and an artificial neural network system using different types of data. For both algorithms.

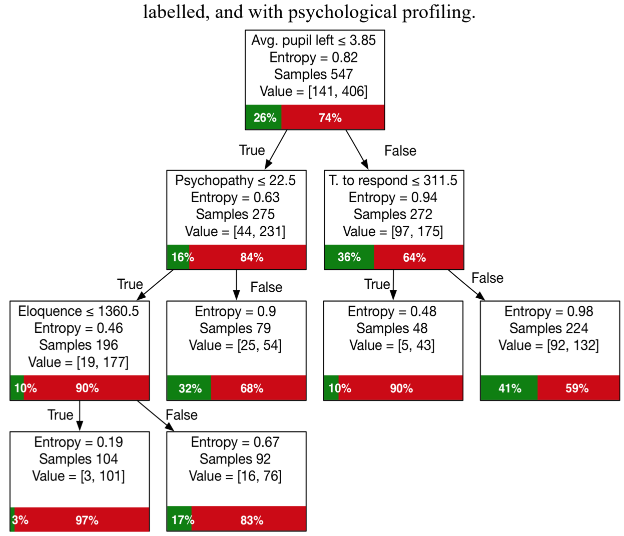
TABLE III. PREDICTION ACCURACY OF A DECISION TREE AND AN ARTIFICIAL NEURAL NETWORK SYSTEM USING DIFFERENT TYPES OF DATA

Features	Prediction accuracy	
	Decision Tree	***MLP***
	D1	D2
Raw	72.4%	64.1%
Raw + psychological profile	69.7%	66.3%
Normalization	75.0%	63.6%
Normalization + psychological profile	70.0%	66.9%
Precise labeling	74.0%	74.2%
Precise labeling + psychological profile	74.0%	74.4%

Different refining of the data set was explored in order to increase accuracy.

Psychological scores from the pre-questionnaire were added to the input vector X. These psychological features could potentially help the models to capture specific behaviors in participants and help to extract psychological trait related to their attitude to lying. The new input vector was 23x1 with 12 psychological scores. A individual adjustment inspired by [26], [34], (Normalization in Table II) was made on temporal speech features. Response time and eloquence were normalized with the baseline data of the experiment to remove speaking differences of participants as suggested by [10].

Figure 5. Most accurate decision tree from cross-validation, precisely labelled, and with psychological profiling.



The last improvement of the dataset was inspired from the analysis of the experiment's videos, it could be observed that some participants chose to deceive using avoidance (*e.g.,* "I don't know") or telling a true statement followed by a lie. These different strategies introduced more variability in the dataset producing ambiguity in the data labeling. A new labeling of the dataset was made depending the different strategies to lie. The lie class was refined by manual labeling using video of the experiments so to identify clear lie. The refining of the labels resulted into an increase of accuracy, for both of classification problems (Table III).

The MLP, and the decision tree classifier algorithms achieved respectively a maximum accuracy of 75% and 74%, supporting the possibility to generalize to detect lie on unprecedented participants. However, the best accuracy was achieved with the new labeling of the dataset. It implies that avoidance and more complex deception strategies are considered as non-lie.

The results achieved with this preliminary work suggests the possibility to train classification models using eyes and speech temporal features to detect lies and true answers. However, in order to develop a more robust lie detection system, it will be helpful to enrich the dataset with more precise labeling. Switching from a binary classification problem to a multi-class classification problem with different labels of deception may be useful to detect and adapt to diverse deceptive strategies.

V. Discussion

Deceit detection is an important skill that an autonomous system can leverage on to support professionals in numerous working context. This research demonstrates that the same set of behavioral cues studied in previous human literature [7], [10], [17]–[19] can be used to detect lies and more importantly can be extracted with non-invasive measures. Moreover, there are no main differences between an interrogation performed by a human and a robot.

This paper analyzes some variables as predictor to a deceptive behavior, which are: (i) time to respond (from the moment the investigator finished the question till the witness started replying); (ii) eloquence time (time the witness spends replying to the question); (iii) number of saccades; (iv) number of fixations; (vi) number of blinks; (vii) left and right pupil dilatations - max, min and average.

The novelty of this work lies on the comparison between participants' behavior when the investigator was a robot or a human and on the development of a machine learning model to detect lies that robots can use.

Some results on the comparison between human and robot, show that the time participants spent talking to the robot or human investigator while telling the truth or lying, did not change significantly (Figure 4). Nevertheless, there is an interesting tendency to show a different behavior in response to the open-ended questions between the robot and human investigators. While telling the truth, participants tend to spend more time talking to the robot compared to the human. This could suggest that the lack of non-verbal feedback of the robot to the participants could evoke in them the need to answer more fluently the questions. On the other side, participants tended to spend more time lying to the human counterpart rather than to the robot. A possible explanation could be that the difficulty to lie to a person is still greater than lying to a robot, as indicated also by the subjective evaluations provided by participants.

These results reflect the possibility that a robot interviewer with anthropomorphic appearance elicits different response during deception act. Although the evidence should be confirmed with more extensive study, this provides preliminary results that robot physical presence might play an important role in the context of interviews.

The study also shades light on the opportunity that humanoid interviewers might be a valid support in professional activities where deceits can undermine the expected goal. The given support could reduce the burden on experienced professional working in demanding contexts such as security, education and healthcare domains.

In fact, the results shown in this paper confirm that a robot could autonomously determine whether an individual is lying or not. In other words, the measurement of the significant features that allow the detection of deception could be in perspective performed by an autonomous robotic platform.

Through experience with participants, the robot can be trained to learn a model that could be reused in successive interactions to detect whether an individual is lying or not.

On top of the favorable results obtained, it is encouraging to improve the experimental protocol to get cleaner data that will improve the accuracy when some individuals decide to deceive avoiding lies or denying the act they had witnessed. For instance, in literature it has been demonstrated that a monetary prize can be used as an incentive to lie. This might constitute a significant drive towards the increase in the effort associated with the deception act. By consequence this could develop into stronger cognitive load associated with the deceit and hence exhibit more evident behavioral cues [10]. However, the results obtained from the autonomous deceit detection from the current data seem already a promising outcome; while indicating a plausible path that would endow a humanoid robot with novel interaction strategies also in relation to deception.

VI. Conclusion

This paper demonstrates the possibility to train a deception detection system with data gathered from interactions with a human and the iCub robotic platform. The study demonstrated the feasibility of the approach and the limitations of the methods adopted. Future work will be focused on refining the experimental methodology, in particular with the aim of better motivating participants to deceive, for instance by providing monetary rewards in case of successful deception of the interrogators.

Another interesting point to explore further will be the effects of the appearance of robots on participants' deceptive behaviors (*e.g.,* virtual agent on screen). Finally, a bigger sample size will be crucial to develop a robust deception detection system that could provide an increased accuracy.

APPENDIX

The location of the videos and the description of the questions are attached after the references.

REFERENCES

[1] B. M. DePaulo, B. E. Malone, J. J. Lindsay, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception," *Psychological Bulletin*. 2003.

[2] M. Zuckerman, B. M. Depaulo, and R. Rosenthal, "Verbal and nonverbal communication of deception," *Adv. Exp. Soc. Psychol.*, 1981.

[3] S. M. Kassin, "On the psychology of confessions: Does innocence put innocents at risk?," *Am. Psychol.*, 2005.

[4] S. M. Kassin and G. H. Gudjonsson, "The Psychology of Confessions:A Review of the Literature and Issues," *Psychol. Sci. Public Interes.*, 2004.

[5] S. M. Kassin and R. J. Norwick, "Why people waive their Miranda rights: The power of innocence," *Law Hum. Behav.*, 2004.

[6] A. Vrij, S. Mann, and R. P. Fisher, "Information-gathering vs accusatory interview style: Individual differences in respondents' experiences," *Pers. Individ. Dif.*, 2006.

[7] S. Leal and A. Vrij, "Blinking during and after lying," *J. Nonverbal Behav.*, vol. 32, no. 4, pp. 187–194, 2008.

[8] A. K. Webb, C. R. Honts, J. C. Kircher, P. Bernhardt, and A. E. Cook, "Effectiveness of pupil diameter in a probable-lie comparison question test for deception," *Leg. Criminol. Psychol.*, 2009.

[9] A. Vrij and P. A. Granhag, "Eliciting cues to deception and truth: What matters are the questions asked," *J. Appl. Res. Mem. Cogn.*, 2012.

[10] J. J. Walczyk, D. A. Griffith, R. Yates, S. R. Visconte, B. Simoneaux, and L. L. Harris, "Lie Detection BY Inducing Cognitive Load: Eye Movements and Other Cues to the False Answers of 'Witnesses' to Crimes," *Crim. Justice Behav.*, vol. 39, no. 7, p. 887–909., 2012.

[11] A. K. Webb, D. J. Hacker, D. Osher, A. E. Cook, D. J. Woltz, S. Kristjansson, and J. C. Kircher, "Eye movements and pupil size reveal deception in computer administered questionnaires," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009.

[12] A. Vrij, J. Oliveira, A. Hammond, and H. Ehrlichman, "Saccadic eye movement rate as a cue to deceit," *J. Appl. Res. Mem. Cogn.*, 2015.

[13] A. Vrij, S. Leal, P. A. Granhag, S. Mann, R. P. Fisher, J. Hillman, and K. Sperry, "Outsmarting the liars: The benefit of asking unanticipated questions," *Law Hum. Behav.*, 2009.

[14] A. Vrij, P. A. Granhag, S. Mann, and S. Leal, "Outsmarting the liars: Toward a cognitive lie detection approach," *Curr. Dir. Psychol. Sci.*, 2011.

[15] K. Fukuda, "Eye blinks: New indices for the detection of deception," *Int. J. Psychophysiol.*, vol. 40, no. 3, pp. 239–245, 2001.

[16] J. A. Stern, L. C. Walrath, and R. Goldstein, "The Endogenous Eyeblink," *Psychophysiology*, 1984.

[17] R. Goldstein, L. C. Walrath, J. A. Stern, and B. D. Strock, "Blink Activity in a Discrimination Task as a Function of Stimulus Modality and Schedule of Presentation," *Psychophysiology*, 1985.

[18] L. O. Bauer, B. D. Strock, R. Goldstein, J. A. Stern, and L. C. Walrath, "Auditory Discrimination and the Eyeblink," *Psychophysiology*, 1985.

[19] J. Beatty and B. Lucero-Wagoner, "The pupillary system," in *Handbook of psychophysiology (2nd ed.).*, 2000.

[20] D. P. Dionisio, E. Granholm, W. A. Hillix, and W. F. Perrine, "Differentiation of deception using pupillary responses as an index of cognitive processing," *Psychophysiology*, vol. 38, no. 2, pp. 205–211, 2001.

[21] J. J. Walczyk, K. S. Roper, E. Seemann, and A. M. Humphrey, "Cognitive mechanisms underlying to questions: Response time as a cue to deception," *Appl. Cogn. Psychol.*, vol. 17, no. 7, pp. 755–774, 2003.

[22] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati, "The benefits of interactions with physically present robots over video-displayed agents," *Int. J. Soc. Robot.*, vol. 3, no. 1, pp. 41–52, 2011.

[23] A. A. Harrison, M. Hwalek, D. F. Raney, and J. G. Fritz, "Cues to Deception in an Interview Situation," *Soc. Psychol. (Gott).*, 1978.

[24] F. Ferrari and F. Eyssel, "Toward a hybrid society," in *International Conference on Social Robotics*, 2016, pp. 909–918.

[25] L. J. Wood, K. Dautenhahn, H. Lehmann, B. Robins, A. Rainer, and D. S. Syrdal, "Robot-mediated interviews: Does a robotic interviewer impact question difficulty and information recovery?," in *Assistive Technology Research Series*, 2013.

[26] J. J. Walczyk, J. P. Schwartz, R. Clifton, B. Adams, M. Wei, and P. Zha, "Lying person-to-person about life events: A cognitive framework for lie detection," *Pers. Psychol.*, 2005.

[27] G. B. Flebus, "Versione Italiana dei Big Five Markers di Goldberg," *Univ. di Milano-Bicocca*, 2006.

[28] D. S. Syrdal, K. Dautenhahn, K. Koay, and M. L. Walters, "The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study," *23rd Conv. Soc. Study Artif. Intell. Simul. Behav. AISB*, pp. 109–115, 2009.

[29] C. J. Ferguson and C. Negy, "Development of a brief screening questionnaire for histrionic personality symptoms," *Pers. Individ. Dif.*, 2014.

[30] D. N. Jones and D. L. Paulhus, "Introducing the Short Dark Triad (SD3): A Brief Measure of Dark Personality Traits," *SAGE*, vol. 21, no. 1, pp. 28–41, 2014.

[31] R. L. Solso, M. K. MacLin, and O. H. MacLin, "Cognitive psychology (7th ed.).," *Cognitive psychology (7th ed.).* 2005.

[32] A. E. Cook, D. J. Hacker, A. K. Webb, D. Osher, S. D. Kristjansson, D. J. Woltz, and J. C. Kircher, "Lyin' eyes: Ocular-motor measures of reading reveal deception," *J. Exp. Psychol. Appl.*, 2012.

[33] T. Qin, J. Burgoon, and J. F. Nunamaker, "An exploratory study on promising cues in deception detection and application of decision tree," *Proc. 37th Annu. Hawaii Int. Conf. Syst. Sci.*, 2004.

[34] D. T. Lykken, *A tremor in the blood: Uses and abuses of the lie detector.* Plenum Press, 1998.

Appendix

A. Videos

Below the links of the videos - they have been edited in order to make them shorter.
Video 1: https://youtu.be/w5sIJhSWw5o
Video 2: https://youtu.be/NGsoqGoUONg

B. General questions - bold questions are verifiable truth.

1. **What is the name of Berlusconi?**
2. **Is Christmas in February?**
3. **Is it possible for a person to get burnt while using an oven?**
4. What is your surname?
5. How old are you?
6. **In which city is located the Coliseum?**
7. Which is your gender?
8. Are you a freshman?
9. **What is the capital of the Italian region of Liguria?**
10. **Is Italy in Europe?**
11. What is the name of your mother?
12. When did you graduated?
13. What is your nationality?
14. Can an oven be hot?
15. **Which day is Christmas?**
16. Are you born before the year 1979?
17. **Are you now in Genova?**
18. Are you a student?
19. **Is the Coliseum located in Bari?**
20. **What language is spoken in the USA?**

B. Video 1, 2 questions - bold questions are short type

Video 1

1. **Were there two or three people?**
2. What was the criminal act?
3. **Was there a person of color?**
4. What is the age of the implicated?
5. **Were they nervous?**
6. Can you describe me the location?
7. **Was there any violence?**
8. Did they wear something to hide their identity?
9. **Were they dressed in elegant clothing?**
10. How many people were during the robbery?
11. **Was there a woman?**
12. How did they hide loot?
13. **Was it crowded?**
14. How did they communicate?
15. **Did they interact with someone?**
16. What did they steal?
17. **Was someone else in the shop?**
18. Where did it happen?
19. **Did they wear summer clothes?**
20. Did they have any particular aspect?

Video 2

1. **Was the criminal male?**
2. What was the criminal act?
3. **Was the person of color?**
4. Did the criminal wear something on the head?
5. **Was someone in the shop?**
6. Where did it happen?
7. **Was the criminal violent?**
8. What was approximatively the age of the criminal?
9. **Was the criminal dressed in elegant clothing?**
10. How did the criminal hide the loot?
11. **Was the criminal female?**
12. How many people were in the shop during the robbery?
13. **Was it committed by a young person?**
14. Was the criminal carrying something in the hands?
15. **Did the criminal interact with someone?**
16. What was robbed?
17. **Did the criminal have a backpack?**
18. Where was the cashier?
19. **Was the criminal wearing sporty clothes?**
20. Did the criminal had any particular sign?