Proc. of the 2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE)
14-15 April 2020, Istanbul, Turkey

Using Eye-tracking to Study the Authenticity of Images Produced by Generative Adversarial Networks

Nicholas Caporusso
Department of Computer Science
Northern Kentucky University
Highland Heights, United States
caporusson1@nku.edu

Kelei Zhang
Department of Informatics
Fort Hays State University
Hays, United States
k_zhang4@fhsu.edu

Gordon Carlson
Department of Communication Studies
Fort Hays State University
Hays, United States
gscarlson@fhsu.edu

***Abstract*—Nowadays, Machine Learning algorithms, such as Generative Adversarial Networks (GANs), enable generating content, and especially images, featuring people, objects, or landscapes, with unprecedented levels of accuracy and fidelity. As a result, it is becoming challenging for a viewer to distinguish a picture of a fake profile from one that has a real human in it. In this paper, we present the results of an experimental study in which we investigated the perception of images produced by GANs. Specifically, we focused on the individuals' ability to discriminate between fake and real profiles. Furthermore, we utilized eye-tracking technology to identify the presence of patterns in subjects' gaze, which, in turn, can be useful to optimize the output of GANs and, simultaneously, provide insight on the underlying cognitive dynamics.**

Keywords—Generative Adversarial Networks, Eye tracking, Cybersecurity, Faceforensics.

I. Introduction

In the last years, the introduction of novel Machine Learning architectures and algorithms, such as Generative Adversarial Networks (GANs) [1] [2], resulted in Artificial Intelligence (AI) systems that are able to generate new content, in addition to classifying and clustering existing data. Specifically, GANs are designed to learn the features of the material in a source dataset (e.g., images, speech, and text) and produce new and realistic content that reproduces the features with an unprecedented level of fidelity and accuracy. To this end, their architecture consists of two neural networks competing against one another: a generator network creates new content based on features and parameters learned from the input dataset, whereas a discriminator network evaluates the output to ensure that its level of realism is consistent with the training set. As a result, GANs can generate high-quality, realistic images. For instance, they can be utilized to produce pictures of non-existing individuals, original artwork that mimics the style of an artist, as well as sophisticated landscapes and objects, which a human viewer finds very difficult to identify as fake.

Studying individuals' perception of content generated with GANs and their behavior in the interaction with it is crucial for improving the fidelity of algorithms and for designing more sophisticated ML systems [3]. Simultaneously, as AI will be able to produce increasingly realistic replicas of user-generated content, it is crucial to identify mechanisms for ensuring users' security as well as interventions that enable educating them against potential malicious applications of this technology.

II. Related Work

Several approaches can be utilized to design algorithms that identify and label content generated with GANs [4][5]. However, the focus of our work is to investigate individuals' ability to recognize images produced with ML and to evaluate the human factors associated with perception of fake content. In [6], the authors discuss how the presence of visual artifacts caused by noise can expose deepfakes and face manipulations. However, in a study that investigated participants' ability to distinguish between AI-generated and real pictures [7], subjects reached approximately 60% accuracy in evaluating content produced by GAN, which demonstrates that the overall quality of the content is high enough to compensate for image aberrations. Specifically, the study found that younger subjects (<45 years) reached higher recognition rates. However, as the experiment was distributed over the Internet, there may be several factors, including heterogeneous contexts, a small number of images included in each session, presence of training effect, individual's ability, or higher familiarity with technology. Therefore, several groups are using different types of physiological signals to measure users' responses to AI-generated images to gain a better understanding of perception dynamics and to identify the components that determine users' ability to distinguish fake and authentic images [8]. In [9], the authors presented a study that analyzed eye blinking as a response to stimuli represented by real pictures and photos generated by ML algorithms. Conversely, the authors of [10] evaluated the possibility of using changes in the electrical activity of the brain as a predictor to improve content generation.

III. Experimental Study

Our work focuses on studying the perception of content generated by Machine Learning algorithms, to identify underlying dynamics that can result in optimized generative systems or can be utilized to counteract potential harmful application of this technology. To this end, we realized an experimental study that aimed at evaluating whether individuals are able to distinguish fake images from real pictures and at analyzing the presence of any common patterns in subjects' gaze that can predict a specific type of content.

A. Participants

Although 54 subjects participated in the study, after the removal of data from 6 individuals who experienced difficulties in completing the trials, in our study, we considered data from 48 participants only. The group consisted of 16 females (33.33%), 20 males (41.67%), and 12 individuals (25%) who preferred not to specify their gender. Participants' age ranged from 18 to 54, with an average of 31±11. Specifically, 19 (39.58%) were in the 18-24 age range, 12 (25%) aged 25-34, 9 (18.75%) were in the 35-44 bracket, and 8 (16.67%) were 45-54. We primarily recruited Gen Y/Millennials (i.e., born between 1980-1994) and Gen Z (i.e., born between 1995 and 2015), because younger individuals are inherently more familiar with Social Media and, thus, more proficient with profile pictures albeit simultaneously being more exposed to the threats of fake profiles. 29 subjects (60.42%) were White or Caucasian, 11 (22.92%) were African American, 4 (8.33%) were Asian, and 4 (8.33%) were Hispanic or Latino. Furthermore, we acquired participants' daily computer use: 2 of them (4.17%) reported less than 1 hour per day, 4 individuals (8.33%) had 1-2 hours of use, 21 (43.75%) used the computer for 2-4 hours per day, 11 (22.92%) had a daily computer use of 4-8 hours, and 10 (20.83%) used the computer for more than 8 hours per day.

B. Materials and Methods

In our study, we utilized a commercial eye-tracking device (Tobii Eye Tracker 4C), which was chosen for its accuracy, compatibility, form factor, and availability of software libraries. It incorporates an array of infrared sensors which, placed at the bottom of the monitor in front of the user, acquire gaze from both eyes at a sampling frequency of 90 Hz. Then, the device converts input signals into a set of X and Y coordinates that represent the point where the user is currently looking at. The acquisition system includes a calibration routine that is designed for adjusting the alignment (both size and position) between the space captured by the eye-tracking device and the monitor. As a result, this improves the resulting accuracy of the area of the screen where the user is looking at. The device and its software were installed on a laptop computer with a screen resolution of 1920×1080 pixels, which was employed to realize the experiment and collect the data. An external USB keyboard was utilized to make it convenient for subjects to type their input without changing their position and, thus, keeping the distance and alignment with respect to the eye-tracking device.

Also, we developed a custom data acquisition software (written in C#) with a two-fold purpose: (1) show the stimuli images (i.e., head shots) and (2) record subjects' responses (i.e., their classification of the stimulus as AI-generated or Real) and gaze movements. To this end, the experimental software consisted of an output component utilized to display a full-screen window showing a sequence of images chosen at random from the pool of AI-generated and real pictures available in the NVIDIA StyleGAN dataset and generated using the Flickr-Faces-HQ Dataset (FFHQ) [2]. Although a new version of StyleGAN is available, we utilized the original dataset, because it contains imperfections and noise, which, in turn, provide subjects with visual cues that may affect participants' attention. The pictures were selected with a ratio of AI-generated images and real images of approximately 3:1. The size of each picture was set to 940×940 pixels to make it clearly visible to the user and to create an adequate distance from the border of the display, in order to compensate for calibration and accuracy errors of the eye-tracking device. Simultaneously, the data collection software acquired the subject's gaze at a sampling rate of 50Hz and recorded their evaluation of the picture.

At the beginning of the experiment, participants were invited to a room where they were seated in front of the experimental equipment. After collecting their demographic information and executing the calibration routine, we instructed the subject to position the hands on the USB keyboard in order to be able to comfortably reach the A key and H key with the left and right hand, respectively. At each trial, the experimental software displayed a different image: the participant had five seconds to make a judgment about the picture before the trial timed out, and the experimental software showed the next one (after a 2-second interval). They could input their choice by pressing the A and H keys, which labeled the image as AI-generated or as real, respectively. The experimental software did not give any feedback to the user in regard to the correctness of their response. We did not establish a limit in terms of the number of trials. As a result, the session continued until the participant pressed the ESC key.

IV. Results and Discussion

The 48 participants considered in our analysis realized a total of 2120 trials (extracted from an initial pool of 54 subjects and trials, as discussed earlier) with a minimum of 13 trials 2143 and a maximum of 103 trials per participant, with an average of 44.16±22.48 trials per participant. Specifically, 8 participants (16.67%) realized less than 20 trials, 14 (29.17%) realized 21-40 trials, 15 (31.25%) realized 41-60 trials, 8 (16.67%) realized 61-80 trials, and 3 (6.25%) realized more than 80 trials. Although having all subjects realize the same number of trials would have allowed a larger and more consistent dataset, our primary objective was to evaluate individuals' ability to recognize fake profiles based on the first impact, rather than studying their performance over time as a result of a training process.

Although the ratio between AI-generated images and real images was set to be approximately 3:1, files were selected at random by the data collection tool at each trial. As a result, the actual percentage of AI-generated images was 73.35%±8.16%, with a minimum and a maximum of AI-generated of 46.67% and 89.47%, respectively. Specifically, the percentage of AI-generated images was between 40% and 50% in 1 trial (2.08%), between 50% and 60% in 1 trial (2.08%), 60%-70% in 10 trials (20.83%), 70%-80% in 29 trials (60.42%), and 80%-90% in 7 trials (14.58%). This allowed us to identify and exclude sessions (or sequences of trials) in which users were consistently evaluating images as belonging to one class, only, instead of making an actual judgment.

Users were able to identify the pictures as in 2065 trials (97.4%), whereas in 55 cases (2.6%), the trial timed out before subjects could make a judgment. The average number of timed out trials was 1.14±1.45 (i.e., 3.08%±4.09%), which demonstrates that all users were able to complete the task in most cases. Given the purpose of our study, in our analysis, we only considered trials that did not time out.

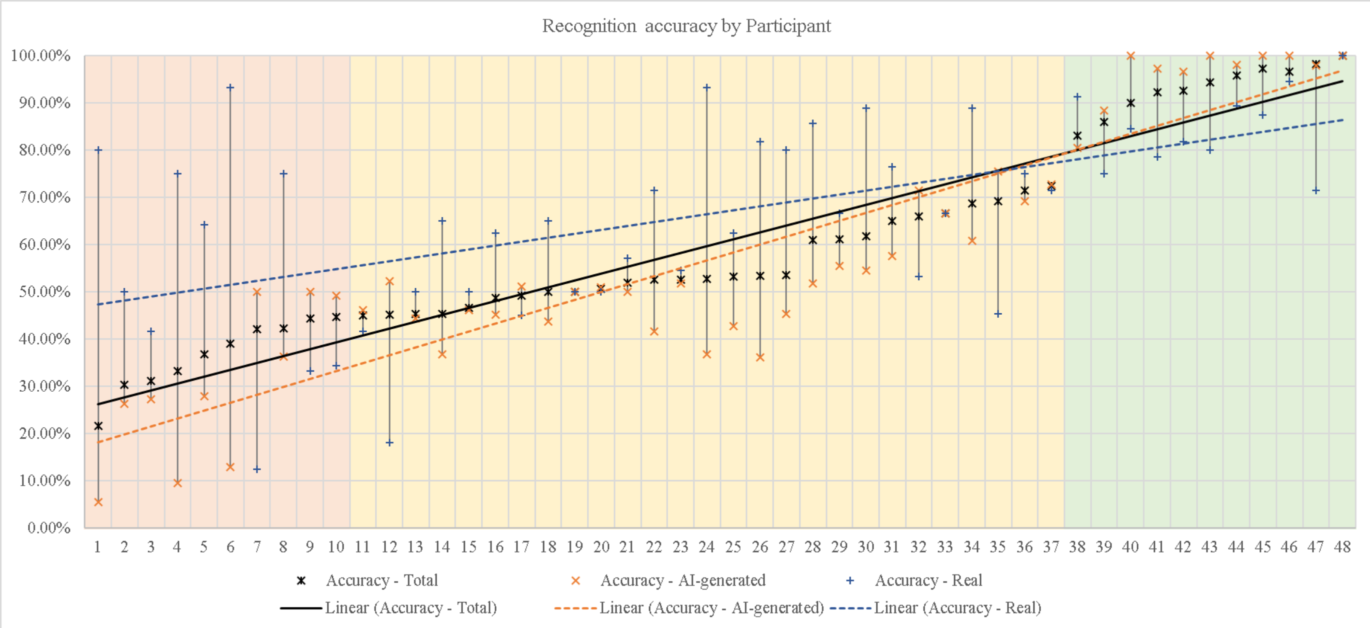

Fig. 1. Participants' results sorted by average recognition rate, calculated as the recognition rate for AI-generated images (AI-RR), Real images (RR-R), and as their compound value (RR-Total).

A. Recognition Accuracy

Overall, recognition accuracy ranged from 21.74% to 100%, with an average of 60.49%±21.10%. For AI images, the recognition rate ranged from 5.56% to 100%, with an average of 57.54%±25.80%. For real pictures, the recognition rate ranged from 12.50% to 100%, with an average of 68.22%±21.37%. AI-generated images had a Precision of 59.84% and a Recall of 84.01%. Conversely, real images had a Precision of 68.32% and a Recall of 37.95%. As AI-generated images had lower Precision (-8.48%) but higher Recall (+46.06%) than real images, they achieved a greater F-score (i.e., 69.90%, compared to 48.79% in real images). Therefore, subjects had better performances with real profiles, though they had more reliable results with AI-generated pictures.

As shown in Figure 1, which represents the data sorted by accuracy (regardless of the number of trials), three groups can be identified based on the overall recognition rate: low accuracy (recognition rate less than 45%), medium accuracy (recognition rate between 45% and 80%), and high accuracy (recognition rate greater than 80%). Specifically, 10 subjects (20.83%) achieved an overall accuracy ranging from 21.74% to 44.66% (average of 36.61%±7.42%), 27 subjects (56.25%) completed the experiment with a recognition rate between 45.10% and 72.41% (55.95%±8.85% on average), and 11 subjects (22.91%) achieved an accuracy ranging from 83.16% to 100% (average of 93.33%±5.23%). Specifically, in the first group (low accuracy), the average recognition rate of AI-generated images is 29.51%±16.82%, whereas the average accuracy in the case of real images is higher (i.e., 55.95%±25.55%). In the second group (medium accuracy), the average recognition rate of AI-generated images is 52.14%±11.10% compared to an average accuracy of 63.58%±17.41% for real images; finally, in the third group (high accuracy) obtained an average accuracy of 96.28%±6.20% with AI-generated images, higher than the average accuracy of real images (84.94%±8.62%). Therefore, the data show that two different dynamics occurring in the different groups: first, users who achieved low and medium overall accuracy are better at recognizing real images rather than AI-generated images, whereas users who have a higher overall accuracy obtained better results with AI-generated images rather than real images; also, as the overall accuracy increases, in addition to a lower variance between the recognition rate of AI-generated and real images within each of the three groups (i.e., 8.73%, 6.31%, and 2.42%, respectively), the difference between the average accuracy in recognizing AI-generated images and real images decreases (26.44% in the first group, 11.44% in the second group, and 11.34% in the third group). Table 1 represents these dynamics in the form of Precision, Recall, and F-score. The difference in terms of accuracy in the recognition of AI-generated images and real images within the first group (low accuracy) is not significant (p=0.693) due to the high variance of data. On the contrary, the difference in recognition rate in the medium accuracy and high accuracy groups are statistically significant (p=0.010) and highly statistically significant (p=0.004), respectively.

Demographic characteristics, such as age, gender, and ethnicity, show almost no correlation (i.e., r values between -0.2 and 0.2) with the ability to recognize images and with the time required to classify an image (i.e., r values between -0.2 and 0.2). Also, the average daily computer use does not have an impact on subjects' accuracy (r values between 0 and 0.2). Therefore, by looking at quantitative numeric indicators about performance only, it is quite impossible to identify any significant patterns in subjects' ability to distinguish AI-generated and real head shots.

TABLE I. PRECISION, RECALL, AND F-SCORE

Accuracy Score	AI-generated			Real		
	P	*R*	*F1*	*P*	*R*	*F1*
Low	75.52%	59.74%	65.92%	32.61%	47.51%	38.67%
Medium	73.14%	69.66%	71.35%	56.44%	60.58%	58.43%
High	75.74%	92.27%	85.17%	97.19%	75.22%	84.80%

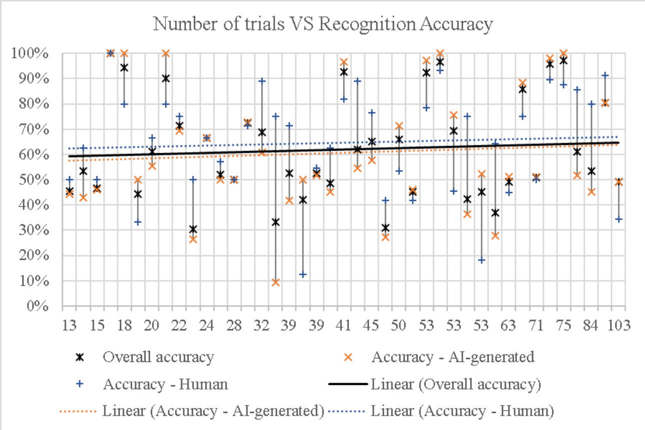

Fig. 2. Success rate (Y-axis) with respect to the number of trials (X-axis). The Figure shows the negligible impact of training on subjects' performances.

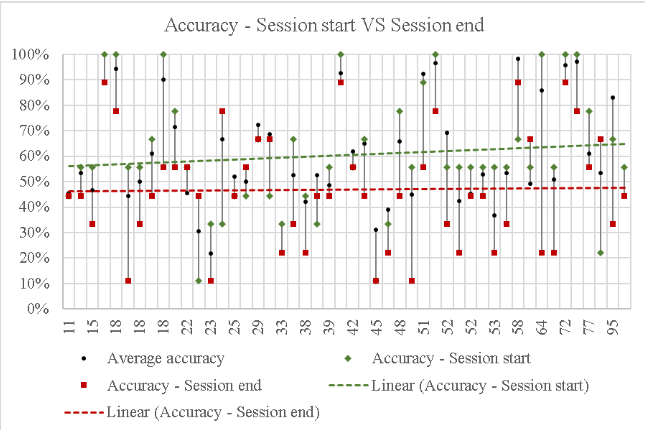

Fig. 3. Accuracy (Y-axis) at the beginning and at the end of the session. The Figure shows that users were performing better at the beginning of the task and that their performance decreases over time. This is consistent across all trials, regardless of their duration (X-axis).

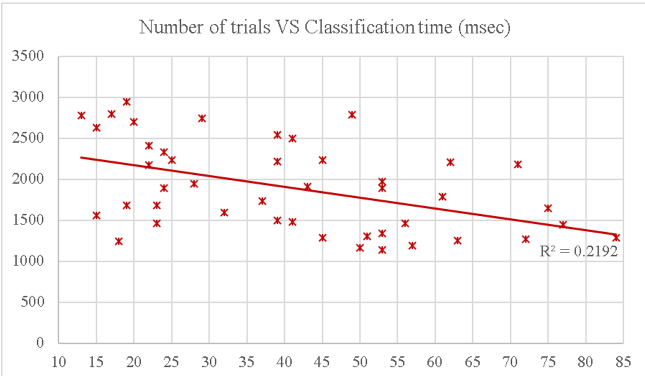

Fig. 4. Classification time (Y-axis) and number of trials (X-axis). The Figure shows the presence of a training effect on the time employed by subjects to classify a picture, though this does not imply a correct recognition.

B. Analysis of Training Effect

Furthermore, we evaluated the presence of any training effect occurring throughout the duration of each session. As shown in Figure 2, the positive correlation between the number of trials in a session, and its overall accuracy is minimal (r=0.157). However, there is no statistical difference between recognition rates in subjects who realized less than 40 trials or longer sessions.

Furthermore, we analyzed 25 sessions (52.08%) of subjects that realized at least 40 trials, and we compared the performance in the first ten trials to the last ten trials. As shown in Figure 3, surprisingly, accuracy decreased in most cases at a statistically significant level (p=0.012): the average recognition rate dropped from 64.89%±23.93% to 41.11%±24.49%. Specifically, 20 individuals (80%) experienced a performance loss between 11.11% and 77.78% (-26.12%±15.83% on average); conversely, 3 subjects (12%) experienced an increase in their recognition rate between 11.11% and 33.33% (+25.92%±12.83% on average), whereas performance did not change in 2 subjects (8%). As the overall recognition accuracy of sessions that had more than 40 trials is slightly better than sessions with fewer trials, the performance loss might be caused by fatigue. This is supported by the data about sessions with less than 40 trials: comparing the performance trend between the first 10 trials and the last 10 trials shows that the performance gains (+57.78%±14.49%) are higher than the losses (-20.37%±8.36%), which may suggest that the ideal session duration for this type of experiment is between 20 and 40 trials. Also, this demonstrates that current GAN algorithms are sophisticated enough to generate high-quality images, which, in turn, render the task of distinguishing an AI-generated picture from a real profile very demanding from a cognitive standpoint.

Also, we analyzed the correlation between the number of trials and the time required by the user to classify an image as AI-generated or human (see Figure 4). Although our data show a moderate negative correlation (r=-0.511) between the two features, faster classification, that is, shorter trial durations do not necessarily result in an image being recognized correctly. On the contrary, there is a slight positive correlation (r=0.218) between trial duration and recognition accuracy. Therefore, the decreasing trend in the time required to classify an image may indicate that as subjects become more familiar with the task, they become more confident and make decisions faster. Additionally, the time employed to type the key associated with their choice may be a factor that is inversely proportional to participants' proficiency with the input method.

C. Eye-tracking

In addition to quantitatively measuring recognition performances, we analyzed data acquired with the eye-tracking device to study gaze and identify any pattern that would suggest the presence of salient cognitive dynamics in participants' interaction with the images. To this end, for each trial, we represented the subject's gaze as a visual map and a heatmap. Visual maps consist of sets of points that represent the sequence of eye movements from the beginning to the end of the trial. They enabled us to: (1) associate a timeline to gaze events and superimpose it on the image, (2) identify dynamic components of subject's attention (e.g., where participants initially focused on), (3) calculate the average speed of the eye during saccades, and (4) evaluate how much of the image subjects explored with their gaze. Also, we utilized heatmaps to superimpose static elements of user's gaze over the image and to: (1) highlight areas where the user spent the most time (i.e., fixations), (2) rank them in terms of importance, and (3) show how much of the image subjects covered with their gaze. Figure 5 shows a heatmap that includes numbers that indicate the sequence of eye movements.

Specifically, we focused our analysis on three indicators, that is, (1) gaze spread (calculated as the average width of eye movements, in pixels, which is a compound measure of the number, length, and speed of saccades), (2) gaze area, that is, the overall width of the area of the image explored by subjects with their gaze (measured as a percentage of the entire size of the picture, which helps identify participants' attention to details beyond facial features), and (3) the facial features that received the most attention (ranked by the number of fixations). Additionally, some areas in several AI-generated images contain errors, such as residual noise and misplaced facial features, caused by the input dataset or by the GAN algorithm itself (e.g., the cutoff threshold of the discriminator). Therefore, we considered them as target areas, and we measured the subjects' ability to spot them by evaluating their fixations. The first three indicators offer insight on subjects' overall attention and on their gaze patterns, whereas the latter can be directly utilized to build a database of noise elements that can be fed into the generator or discriminator algorithms of the GAN so that an ad hoc filter can remove them or discard the generated image, respectively, to improve the fidelity of the output picture.

For the purpose of our analysis, we extracted data from 18 sessions (37.5% of the total) realized by 9 subjects with consistently good performances (accuracy higher than 80%) and 9 with consistently bad performances (accuracy lower than 40%), as detailed in Figure 1. We especially focused on these groups, because the gaze patterns from their sessions are consistent across the trials, and they clearly exhibit distinct features that render the differences clearer from a data analysis and visualization standpoint. The 18 subjects realized an average of 42±25 trials per session, for a total of 753 trials, 400 (53.12%) of which belong to the high accuracy group, and 353 (46.87%) to subjects with consistently low accuracy.

Gaze spread was calculated by evaluating the distance (in pixels) of the coordinates of the eyes from the center point of the picture (see Figure 6). The overall average gaze spread was 191.34±105.37 pixels (18.68%±10.29%). The average gaze spread of the group with high accuracy was 231.43±98.27 pixels (22.60%±9.59% of the width of the image) whereas the low accuracy subjects had an average gaze spread of 145.91±94.21 pixels (14.24%±9.20% of the width of the image). Our data show that subjects who covered a larger distance with their gaze achieved better results. However, on average, trials in the high-performance group had a longer duration (3.10±1.4 seconds) compared to the group with lower recognition rates (2.51±1.1 seconds). Wwe divided the gaze spread by the duration of the trial to normalize its values. As a result, we obtained the average gaze spread velocity, which was 7.45±3.16 pixels/msec and 5.79±3.74 pixels/msec in the high and low accuracy groups, respectively, confirming that individuals who show higher values of gaze spread achieved better results in terms of recognition rate, regardless of the duration of their trials.

Furthermore, we analyzed the gaze area, which represents to which extent each participant explored the picture, calculated as a percentage of the surface of the image covered by the subject's gaze. To this end, we represented gaze as a circle having a radius of 30 pixels (3% of the width of the image). This enabled us to calculate the area where the eyes of the user were focusing on at each fixation. The resulting heatmap representing gaze areas from all subjects is shown in Figure 7. On average, participants explored 115090 ±113659 pixels, that is, 13.0%±12.9% of each image (883.600 pixels). However, subjects in the high accuracy group covered a larger area compared to the low accuracy group, that is, 164485±119883 pixels (18.62%±13.57%) and 59118±73439 pixels (6.69%±8.31%), respectively. The differences are statistically significant at α=0.01. Gaze areas from individual trials are shown in Figure 8, which represents them as bounding boxes, each delimiting subjects' gaze area. Although this representation is not accurate, because vertices are calculated based on the minimum and maximum eye position coordinates, it serves the purpose of achieving a better visualization of the differences in the two groups. Our findings suggest that recognition accuracy does not depend on an individual's demographic characteristics. This contrasts with a previous study, which demonstrated a negative correlation between age and accuracy [7]. However, this might be due to the smaller sample and different age distribution of our subject group. Furthermore, familiarity with and use of the computer does not have an impact on performance. As a result, an individual's recognition accuracy may depend on other factors beyond the purpose of this study. Nevertheless, the analysis of gaze patterns reveals that acquiring more information from the image enables them to make better decisions. Furthermore, the results of our study may be relevant in a cybersecurity context as they suggest that individuals inherently tend to focus mostly on prominent facial features, such as the eyes (56% of fixation time) and the mouth (23% of fixation time) when asked to distinguish a fake profile from a real picture. They are prone to disregarding other areas of the image that may contain components, such as aberrations, that are useful for identifying AI-generated content.

Fig. 5. Heatmaps obtained by superimposing users' fixations on the original image. The Figure shows characteristic gaze patterns, such as scanning the entire face, focusing on facial features (e.g., the eyes or the mouth), exploring the background, or identifying image aberrations.

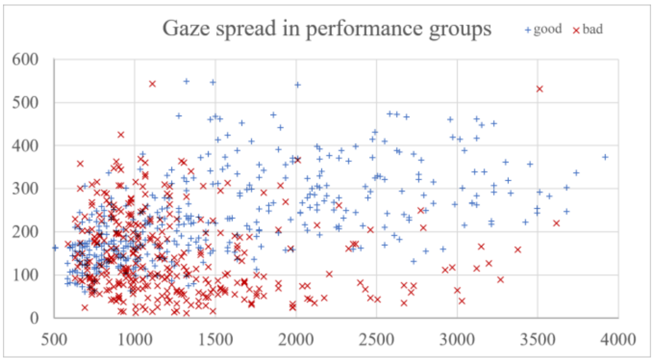

Fig. 6. Gaze spread in pixels (X-axis) and trial duration in milliseconds (Y-axis), in groups with good (blue) and bad (red) performances, calculated as the average space covered by subjects' saccades, which represents the overall speed and width of eye movements. The Figure shows that individuals who explored the image with larger saccades achieve better accuracy than subjects who had the tendency to focus on specific spots.

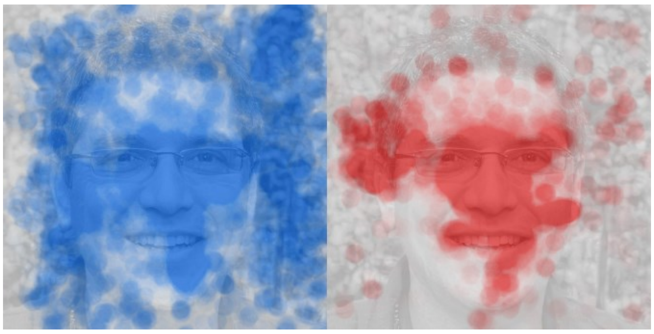

Fig. 7. Heatmap representing areas of the image and facial features that received the most attention based on overall data from the group with high accuracy (left, in blue) and the group with low accuracy (right, in red).

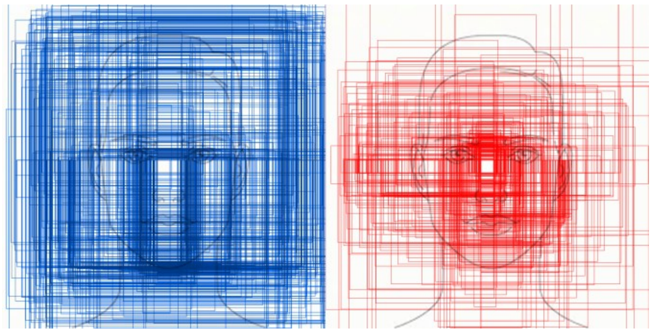

Fig. 8. Bounding boxes representing the gaze area in individual trials from the group with high accuracy (left, in blue) and the group with low accuracy (right, in red).

V. Conclusions and Future Work

In this paper, we presented the results of an experimental study in which we evaluated individuals' accuracy in distinguishing fake profiles generated by GANs and pictures featuring real people. To this end, we provided our participants with sequences of images selected at random and asked them to flag their content as authentic or AI-generated, while simultaneously acquiring their gaze. Subjects achieved approximately 60% accuracy in terms of recognition rate, consistently with other studies in the literature. However, the data from their sessions showed specific features that enabled us to divide them in three distinct groups based on their accuracy. Although participants' performances do not show any correlation with any of the other characteristics taken into consideration in this study (e.g., demographic information, familiarity with computer), by investigating their gaze, we could identify patterns that can serve as a statistically significant predictor of subjects' ability to distinguish AI-generated pictures from images featuring real humans.

As our work might be especially useful in improving the generation of images for healthcare purposes [11], in our future work, we will expand our participant pool, which will enable us to label and rank images based on their level of realism. Also, we will focus on exploring specific perception and cognitive dynamics related to the gaze patterns that have been identified in the study, as well as whether AI-generated content reaches the same level of likeness as real human faces. To this end, we will use both the old and the new versions of the StyleGAN datasets. Furthermore, we will make changes to our experimental protocol in order to provide feedback to the user about their response in each trial to specifically analyze how training dynamics influence user performance and gaze patterns.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[2] Karras, T., Laine, S. and Aila, T., 2019. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4401-4410).

[3] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Nießner, M., 2019. Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1-11).

[4] Yang, X., Li, Y. and Lyu, S., 2019, May. Exposing deep fakes using inconsistent head poses. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8261-8265). IEEE.

[5] McCloskey, S. and Albright, M., 2018. Detecting gan-generated imagery using color cues. arXiv preprint arXiv:1812.08247.

[6] Matern, F., Riess, C. and Stamminger, M., 2019, January. Exploiting visual artifacts to expose deepfakes and face manipulations. In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW) (pp. 83-92). IEEE.

[7] Caporusso, N., Zhang, K., Carlson, G., Jachetta, D., Patchin, D., Romeiser, S., Vaughn, N. and Walters, A., 2019, August. User Discrimination of Content Produced by Generative Adversarial Networks. In International Conference on Human Interaction and Emerging Technologies (pp. 725-730). Springer, Cham.

[8] Ciftci, U.A. and Demir, I., 2019. FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals. arXiv preprint arXiv:1901.02212.

[9] Li, Y., Chang, M.C., Farid, H. and Lyu, S., 2018. In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking. arXiv preprint arXiv:1806.02877.

[10] Palazzo, S., Spampinato, C., Kavasidis, I., Giordano, D. and Shah, M., 2017. Generative adversarial networks conditioned by brain signals. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3410-3418).

[11] V.Bevilacqua, L.Carnimeo, A.Brunetti, A.De Pace, P.Galeandro, G.F.Trotta, N.Caporusso, F.Marino, V.Alberotanza, A.Scardapane, Synthesis of a Neural Network Classifier for Hepatocellular Carcinoma Grading based on triphasic CT images, in Proc. of the International Conf. on Recent Trends in Image Processing & Pattern Recognition, 2016.