Synthetic Media

for

Parallel Hyperrealities

Supervisor:
Vertr.-Prof. Jason M. Reizner

Author:
Zeinab Rahimi

SCAN ME

Bauhaus-Universität Weimar

MA MediaArchitecture Sammersemester 2022

REAL

Or

FAKE

CONTENT :

Introduction

Introduction :

Thanks to the advancement of computational systems and the emergence of large datasets, in recent decades, artificial intelligence knowledge and its applications in various sciences have grown rapidly.

In the past few years, a new type of algorithm called Generative Adversarial Network (GAN) has appeared that allows us to generate artificial faces that are identical to real faces. GANs are a deep-learning-based generative model. It can be described as an approach to generative modeling using deep learning methods, such as convolutional neural networks. The GAN architecture was first described in the 2014 paper by Ian Goodfellow, et al. titled “Generative Adversarial Networks”. [1]

This algorithm is currently widely used in various examples, including the generation of new faces for marketing campaigns, video processing, increasing the resolution of images, as well as in entertainment applications. GAN lets you generate artificial faces that look like real ones

Real or *Fake* **?**

Introduction

The similarity of the images generated by GAN to the real photos has raised many concerns, because it can be used for unethical or illegal activities.

As a result of these concerns, numerous algorithms have been developed by researchers to differentiate between generated and real images. However, by reviewing the literature, we realize that the human ability to understand and recognize generated images remains relatively unexplored.

In this study, we examine how well people are able to recognize, distinguish, and memorize real, fake and edited faces.

SCAN ME

Problem Statement

Problem Statement :

With the availability of generative technologies to all people, there is also the concern of misusing this generated data. Below are some highlights from researchers about the potential problems of this technology:

"Deep learning advances however have also been employed to create software that can cause threats to privacy, democracy and national security. One of those deep learning-powered applications recently emerged is deepfake. Deepfake algorithms can create fake images and videos that humans cannot distinguish them from authentic ones. The proposal of technologies that can automatically detect and assess the integrity of digital visual media is therefore indispensable."[2]

"Computer graphics techniques for image generation are living an era where, day after day, the quality of produced content is impressing even the more skeptical viewer. Although it is a great advance for industries like games and movies, it can become a real problem when the application of such techniques is applied for the production of fake images."[3]

P r o b l e m   S t a t e m e n t

Problem Statement :

"The first deepfake video emerged in 2017 where face of a celebrity was swapped to the face of a porn actor. It is threatening to world security when deepfake methods can be employed to create videos of world leaders with fake speeches for falsification purposes [4–6].

Deepfakes can be abused to cause political or religion tensions between countries, to fool public and affect results in election campaigns, or create chaos in financial markets by creating fake news [7–9].

It can be used to generate fake satellite images of the Earth to contain objects that do not really exist to confuse military analysts, e.g., creating a fake bridge across a river although there is no such a bridge in reality. This can mislead a troop who have been guided to cross the bridge in a battle [10, 11]." [2]

The use of deepfakes can also be positive, such as in visual effects, digital avatars, snapshot filters, reviving lost voices, or updating episodes of movies without reshooting them [12]. Among the many creative or productive impacts that deepfakes can have on photography, video games, virtual reality, movie productions, and entertainment are realistic video dubbings of foreign films, reanimations of historical figures for education, and virtual clothing try-ons for shopping [13, 14]. Despite this, the number of malicious uses of deepfakes greatly outweighs those that are beneficial., therefore finding the truth in digital domain has become increasingly critical. [2]

Problem Statement

To solve this potential problem, there have been numerous methods and tools proposed to detect computer generated photos from real photos. Most of them are based on deep learning. [2]

However, according to the literature review, it has been clear that the human ability to recognize generated images, as well as how humans perceive and communicate emotionally with the generated photos, has not been adequately explored.

In this project we examine how well people are able to **recognize**, **distinguish**, **memorize** and **perceive** real and fake faces.

Zeinab Rahimi

Real or *Fake* ?

Problem Statement

Table 2: Summary of prominent deepfake detection methods

Methods	Classifiers/ Techniques	Key Features	Dealing with	Datasets Used
Eye blinking [114]	LRCN	- Use LRCN to learn the temporal patterns of eye blinking. - Based on the observation that blinking frequency of deepfakes is much smaller than normal.	Videos	Consist of 49 interview and presentation videos, and their corresponding generated deepfakes.
Intra-frame and temporal inconsistencies [112]	CNN and LSTM	CNN is employed to extract frame-level features, which are distributed to LSTM to construct sequence descriptor useful for classification.	Videos	A collection of 600 videos obtained from multiple websites.
Using face warping artifacts [120]	VGG16 [119], ResNet models [118]	Artifacts are discovered using CNN models based on resolution inconsistency between the warped face area and the surrounding context.	Videos	- UADFV [121], containing 49 real videos and 49 fake videos with 32752 frames in total. - DeepfakeTIMIT [69]
MesoNet [102]	CNN	- Two deep networks, i.e. Meso-4 and MesoInception-4 are introduced to examine deepfake videos at the mesoscopic analysis level. - Accuracy obtained on deepfake and FaceForensics datasets are 98% and 95% respectively.	Videos	Two datasets: deepfake one constituted from online videos and the FaceForensics one created by the Face2Face approach [52].
Eye, teach and facial texture [130]	Logistic regression and neural network (NN)	- Exploit facial texture differences, and missing reflections and details in eye and teeth areas of deepfakes. - Logistic regression and NN are used for classifying.	Videos	A video dataset downloaded from YouTube.
Spatio-temporal features with RCN [103]	RCN	Temporal discrepancies across frames are explored using RCN that integrates convolutional network DenseNet [62] and the gated recurrent unit cells [111]	Videos	FaceForensics++ dataset, including 1,000 videos [105].
Spatio-temporal features with LSTM [139]	Convolutional bidirectional recurrent LSTM network	- An XceptionNet CNN is used for facial feature extraction while audio embeddings are obtained by stacking multiple convolution modules. - Two loss functions, i.e. cross-entropy and Kullback-Leibler divergence, are used.	Videos	FaceForensics++ [105] and Celeb-DF (5,639 deepfake videos) [107] datasets and the ASVSpoof 2019 Logical Access audio dataset [140].
Analysis of PRNU [131]	PRNU	- Analysis of noise patterns of light sensitive sensors of digital cameras due to their factory defects. - Explore the differences of PRNU patterns between the authentic and deepfake videos because face swapping is believed to alter the local PRNU patterns.	Videos	Created by the authors, including 10 authentic and 16 deepfake videos using DeepFaceLab [38].
Phoneme-viseme mismatches [141]	CNN	- Exploit the mismatches between the dynamics of the mouth shape, i.e. visemes, with a spoken phoneme. - Focus on sounds associated with the M, B and P phonemes as they require complete mouth closure while deepfakes often incorrectly synthesize it.	Videos	Four in-the-wild lip-sync deepfakes from Instagram and YouTube (www.instagram.com/bill_posters_uk and youtu.be/VWMEDacz3L4) and others are created using synthesis techniques, i.e. Audio-to-Video (A2V) [68] and Text-to-Video (T2V) [142].
Using attribution-based confidence (ABC) metric [143]	ResNet50 model [118], pre-trained on VGGFace2 [144]	- The ABC metric [145] is used to detect deepfake videos without accessing to training data. - ABC values obtained for original videos are greater than 0.94 while those of deepfakes have low ABC values.	Videos	VidTIMIT and two other original datasets obtained from the COHFACE (https://www.idiap.ch/dataset/cohface) and from YouTube. datasets from COHFACE [146] and YouTube are used to generate two deepfake datasets by commercial website https://deepfakesweb.com and another deepfake dataset is DeepfakeTIMIT [147].
Using appearance and behaviour [148]	Rules based on facial and behavioural features.	Temporal, behavioral biometric based on facial expressions and head movements are learned using ResNet-101 [118] while static facial biometric is obtained using VGG [65].	Videos	The world leaders dataset [1], FaceForensics++ [105], Google/Jigsaw deepfake detection dataset [106], DFDC [109] and Celeb-DF [107].
FakeCatcher [149]	CNN	Extract biological signals in portrait videos and use them as an implicit descriptor of authenticity because they are not spatially and temporally well-preserved in deepfakes.	Videos	UADFV [121], FaceForensics [127], FaceForensics++ [105], Celeb-DF [107], and a new dataset of 142 videos, independent of the generative model, resolution, compression, content, and context.
Emotion audio-visual affective cues [150]	Siamese network [86]	Modality and emotion embedding vectors for the face and speech are extracted for deepfake detection.	Videos	DeepfakeTIMIT [147] and DFDC [109].
Head poses [121]	SVM	- Features are extracted using 68 landmarks of the face region. - Use SVM to classify using the extracted features.	Videos/ Images	- UADFV consists of 49 deep fake videos and their respective real videos. - 241 real images and 252 deep fake images from DARPA MediFor GAN Image/Video Challenge.
Capsule-forensics [123]	Capsule networks	- Latent features extracted by VGG-19 network [119] are fed into the capsule network for classification. - A dynamic routing algorithm [125] is used to route the outputs of three convolutional capsules to two output capsules, one for fake and another for real images, through a number of iterations.	Videos/ Images	Four datasets: the Idiap Research Institute replay-attack [126], deepfake face swapping by [102], facial reenactment FaceForensics [127], and fully computer-generated image set using [128].
Preprocessing combined with deep network [74]	DCGAN, WGAN-GP and PGGAN.	- Enhance generalization ability of deep learning models to detect GAN generated images. - Remove low level features of fake images. - Force deep networks to focus more on pixel level similarity between fake and real images to improve generalization ability.	Images	- Real dataset: CelebA-HQ [61], including high quality face images of 1024x1024 resolution. - Fake datasets: generated by DCGAN [88], WGAN-GP [90] and PGGAN [61].
Analyzing convolutional traces [151]	KNN, SVM, and linear discriminant analysis (LDA)	Using expectation-maximization algorithm to extract local features pertaining to convolutional generative process of GAN-based image deepfake generators.	Images	Authentic images from CelebA and corresponding deepfakes are created by five different GANs (group-wise deep whitening-and-coloring transformation GDWCT [152], StarGAN [153], AttGAN [154], StyleGAN [51], StyleGAN2 [155]).
Bag of words and shallow classifiers [78]	SVM, RF, MLP	Extract discriminant features using bag of words method and feed these features into SVM, RF and MLP for binary classification: innocent vs fabricated.	Images	The well-known LFW face database [156], containing 13,223 images with resolution of 250x250.
Pairwise learning [85]	CNN concatenated to CFFN	Two-phase procedure: feature extraction using CFFN based on the Siamese network architecture [86] and classification using CNN.	Images	- Face images: real ones from CelebA [87], and fake ones generated by DCGAN [88], WGAN [89], WGAN-GP [90], least squares GAN [91], and PG-GAN [61]. - General images: real ones from ILSVRC12 [92], and fake ones generated by BIGGAN [93], self-attention GAN [94] and spectral normalization GAN [95].
Defenses against adversarial perturbations in deepfakes [157]	VGG [65] and ResNet [118]	- Introduce adversarial perturbations to enhance deepfakes and fool deepfake detectors. - Improve accuracy of deepfake detectors using Lipschitz regularization and deep image prior techniques.	Images	5,000 real images from CelebA [87] and 5,000 fake images created by the "Few-Shot Face Translation GAN" method [158].
Face X-ray [159]	CNN	- Try to locate the blending boundary between the target and original faces instead of capturing the synthesized artifacts of specific manipulations. - Can be trained without fake images.	Images	FaceForensics++ [105], DeepfakeDetection (DFD) [106], DFDC [109] and Celeb-DF [107].
Using common artifacts of CNN-generated images [160]	ResNet-50 [118] pre-trained with ImageNet [92]	Train the classifier using a large number of fake images generated by a high-performing unconditional GAN model, i.e., PGGAN [61] and evaluate how well the classifier generalizes to other CNN-synthesized images.	Images	A new dataset of CNN-generated images, namely ForenSynths, consisting of synthesized images from 11 models such as StyleGAN [51], super-resolution methods [161] and FaceForensics++ [105].
Using convolutional traces on GAN-based images [162]	KNN, SVM, and LDA	Training the expectation-maximization algorithm [163] to detect and extract discriminative features via a fingerprint that represents the convolutional traces left by GANs during image generation.	Images	A dataset of images generated by ten GAN models, including CycleGAN [164], StarGAN [153], AttGAN [154], GDWCT [152], StyleGAN [51], StyleGAN2 [155], PGGAN [61], FaceForensics++ [105], IMLE [165], and SPADE [42].
Using deep features extracted by CNN [100]	A new CNN model, namely SCnet	The CNN-based SCnet is able to automatically learn high-level forensics features of image data thanks to a hierarchical feature extraction block, which is formed by stacking four convolutional layers.	Images	A dataset of 321,378 face images, created by applying the Glow model [101] to the CelebA face image dataset [87].

Deepfake detection methods:
Source: https://doi.org/10.1016/j.cviu.2022.103525

Method and Design

Method and Design:

The online platform:

An online platform called **RealorFake** was developed to collect the opinions of the respondents (https://realorfake.netlify.app). It is a two-page website based on the Bare- Bootstrap Starter Template, which was customized according to the project's needs and the author's taste. Netlify.com is used to publish the website.

The first page of the website gives introductory information about the project and its purpose. On the second page, there is a survey form. To run the survey, we used Google Forms, an easy-to-use and reliable form.

The online method is chosen because it is practical and accessible. Nowadays, computers and the Internet are part of our daily lives. Therefore, more people will have the opportunity to participate, and each participant will be able to choose when and where to participate based on their own preferences.

Method and Design

The survey form:

The survey form has three sections. The first section is a “linear test” which means the participant will make a decision based only on one picture. In this section, they have to answer two multiple-choice questions. The first question is " Do you think this person is real?”, and the second question is asking if they can tell which part of the photo is making them believe the photo is not real.

In the second section, the goal is to find out how memorable the photos are to the participants and therefore a "comparison test" is done, which means an entire table of all the photos is shown and the participants must choose which one is most memorable to them.

The third section is designed to collect respondents’ feelings and perceptions in terms of "trust". This section is also conducted with a "comparative test", which shows the entire set of photos and asks the question " Which Face do you trust more?".

- **Questions :**
- Do you think this person is real?
- If you think this photo is fake or edited, which part of photo makes you think so?
- Which Face do you remember best(**Memory**)?
- Which Face do you trust more **(Perception**)?

Method and Design

The Photos:

The photos in this project are categorized into three groups of three each. A total of nine photos were shown to respondents. The first group is *real* people, the second group is real photos that are *edited* by software like Photoshop. The third group is photos that are computer generated and the people do not exist in real life!

All pictures were taken from open-source databases. For computer-generated faces, the website “This Person Does Not Exist” (https://thispersondoesnotexist.com/) is used and, for real faces: FFHQ dataset (https://github.com/NVlabs/ffhq-dataset). In choosing the photos, age, gender, and skin color were taken into consideration.

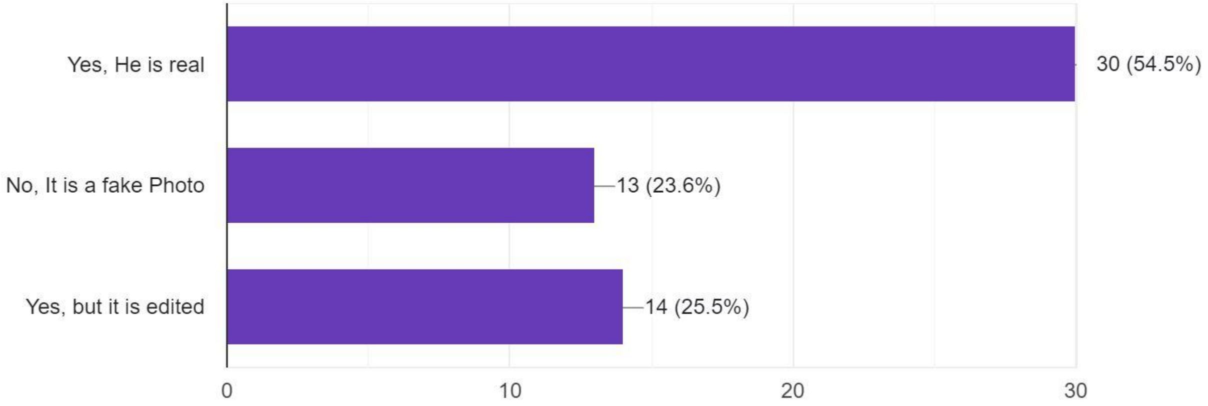
Result

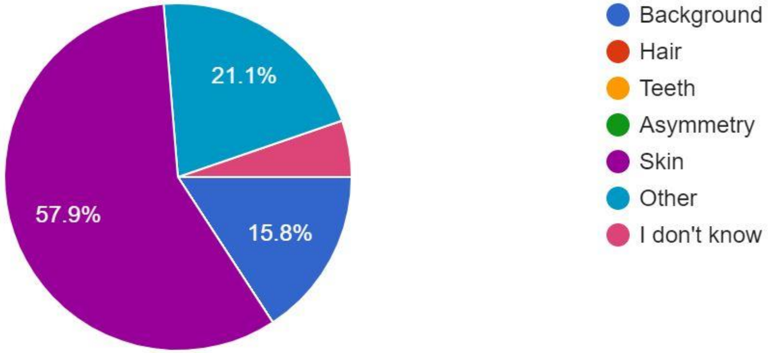
Real **or** ***Fake*** **?**

SCAN ME

Real or *Fake* ?

This Photo is real.



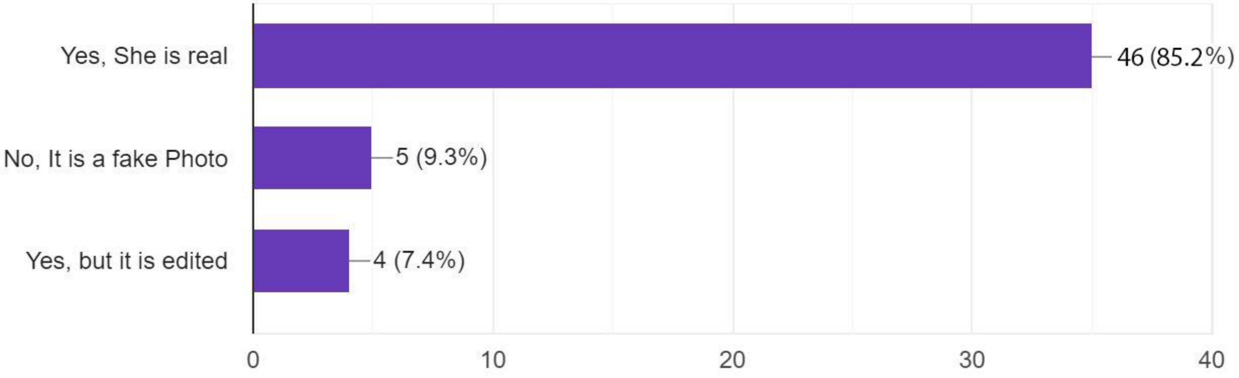


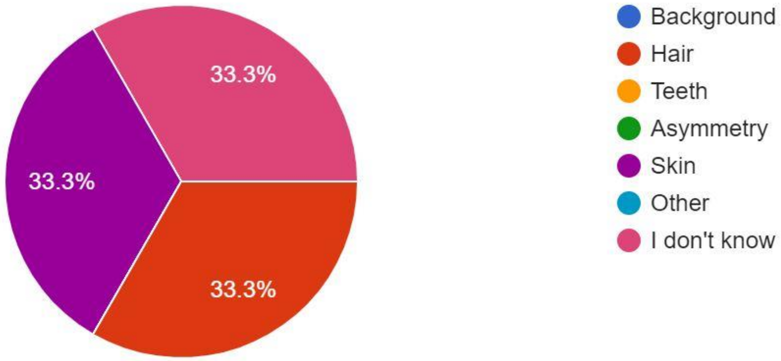
Synthetic Media for Parallel Hyperrealities

Zeinab Rahimi

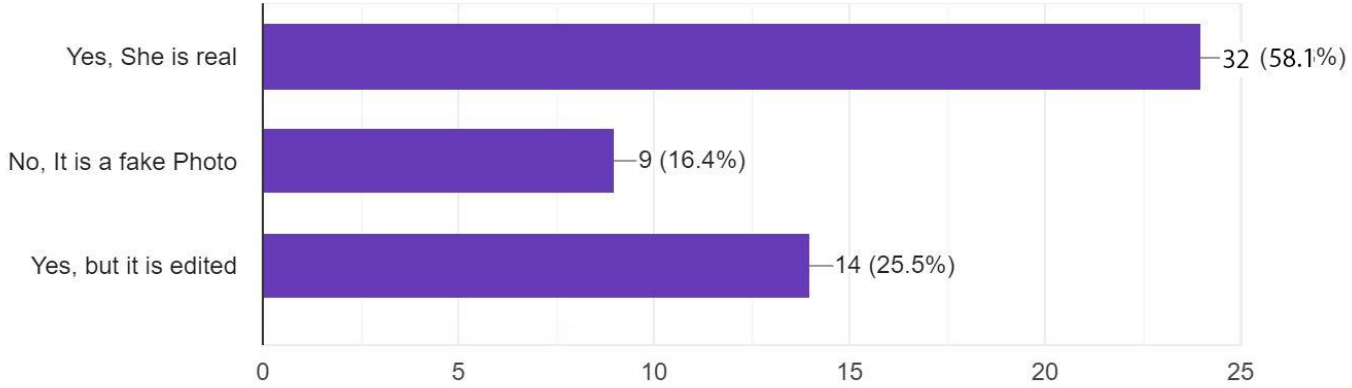
Real or *Fake* ?

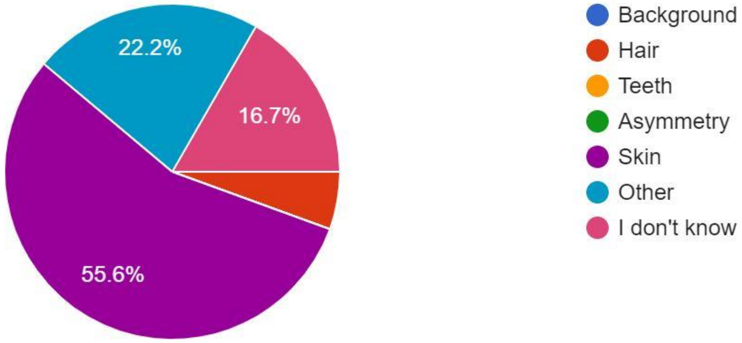
This Photo is real.





Real or *Fake* ?



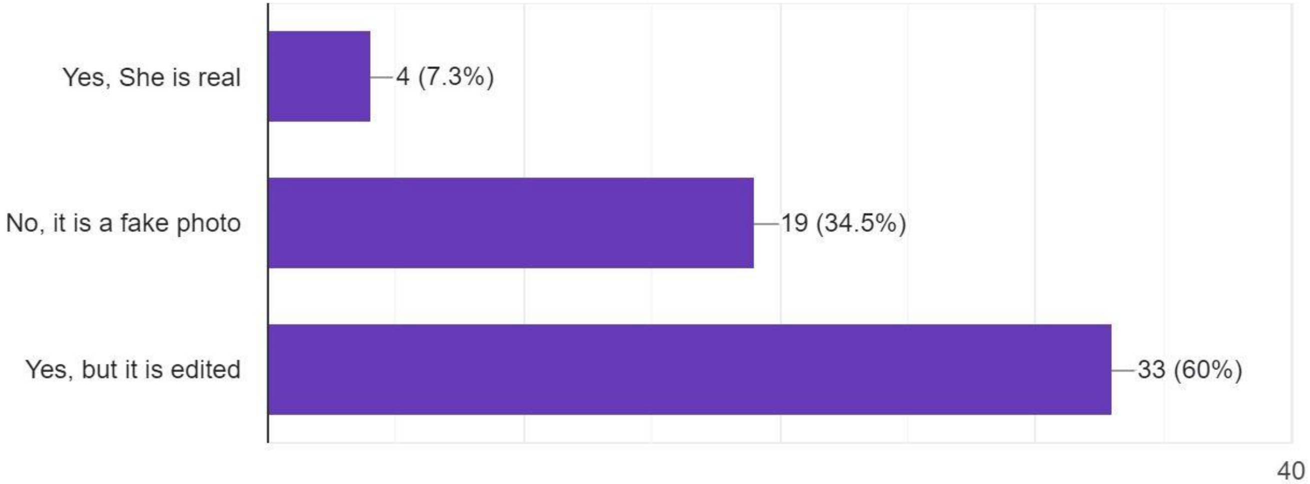


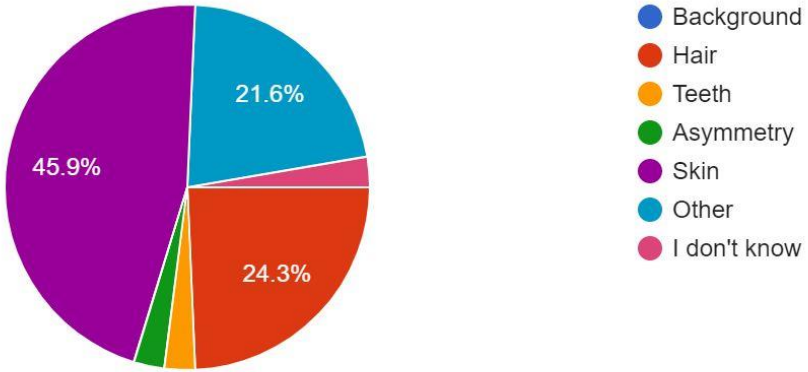
This Photo is real.

Zeinab Rahimi

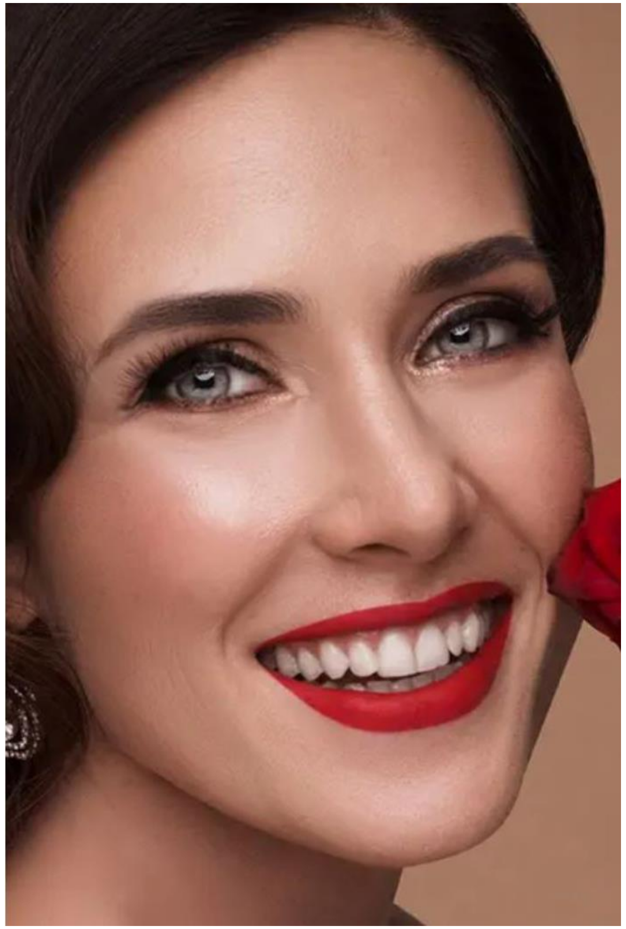
Real or *Fake* ?

This Photo is real but edited.

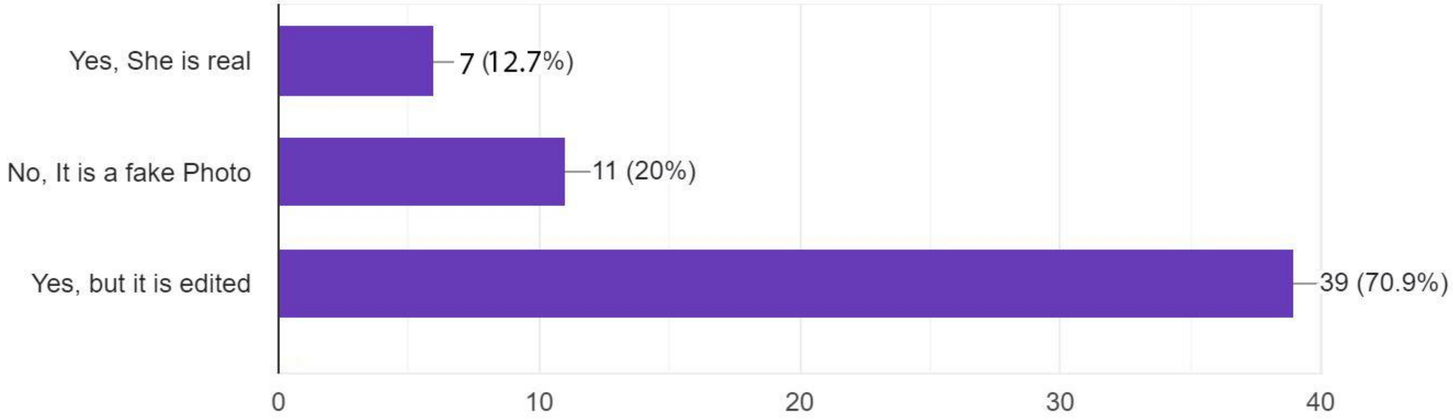


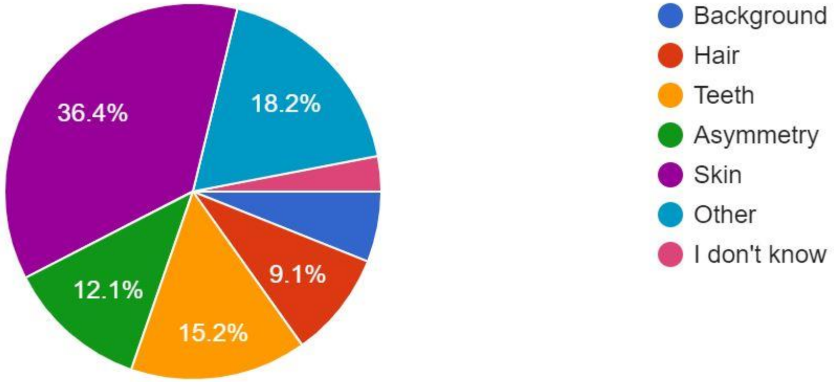


Real or *Fake* ?



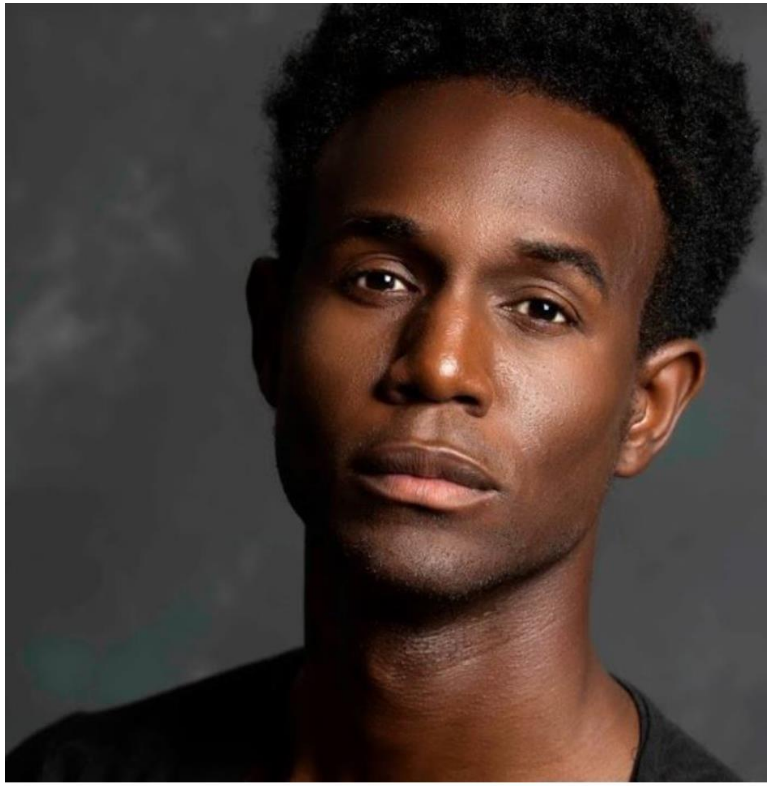
This Photo is real but edited.



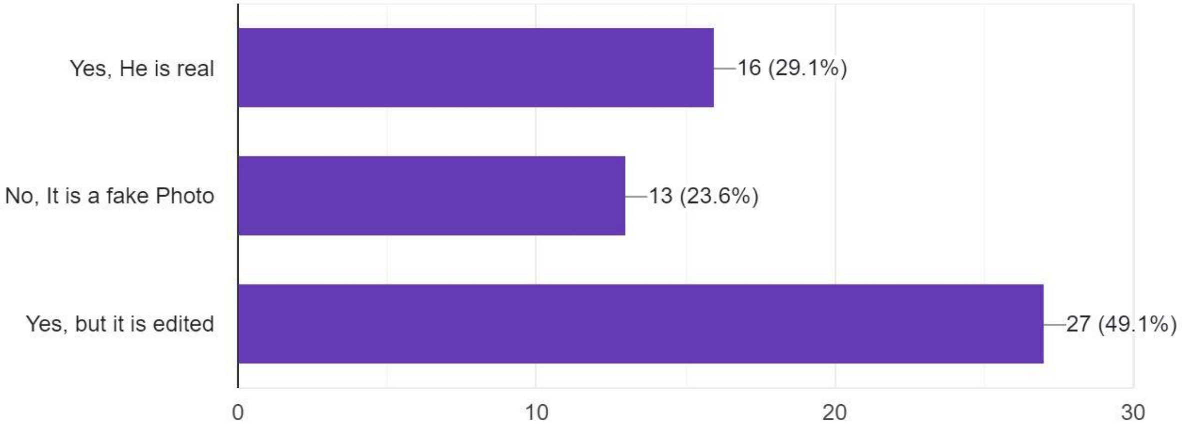


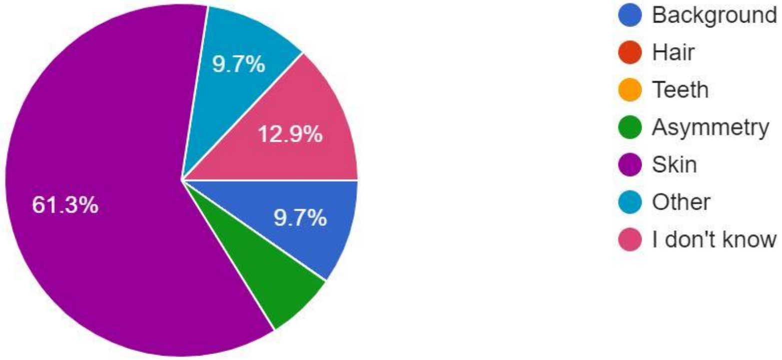
Zeinab Rahimi

Real or *Fake* ?



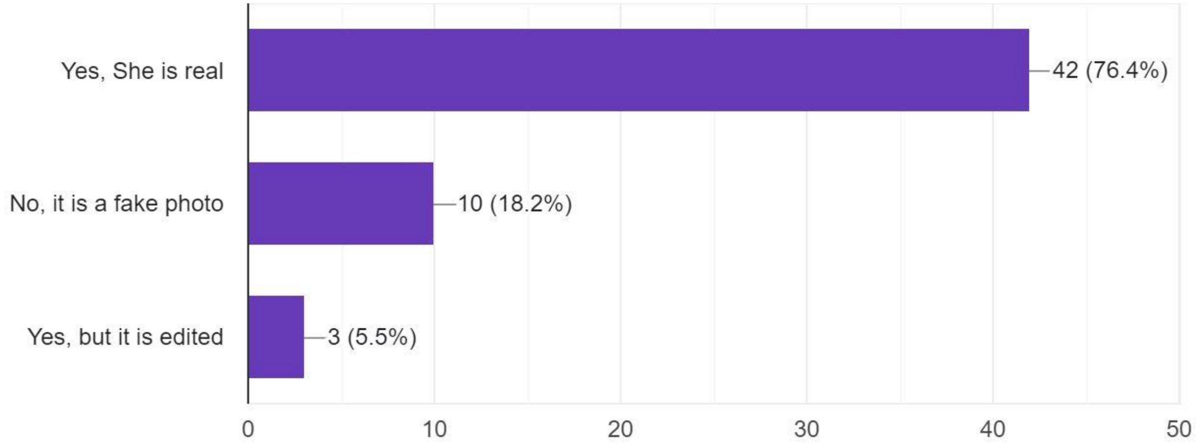
This Photo is real but edited.

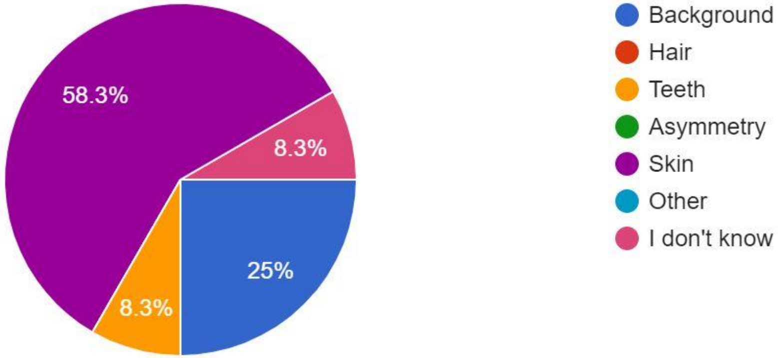




Real or *Fake* ?

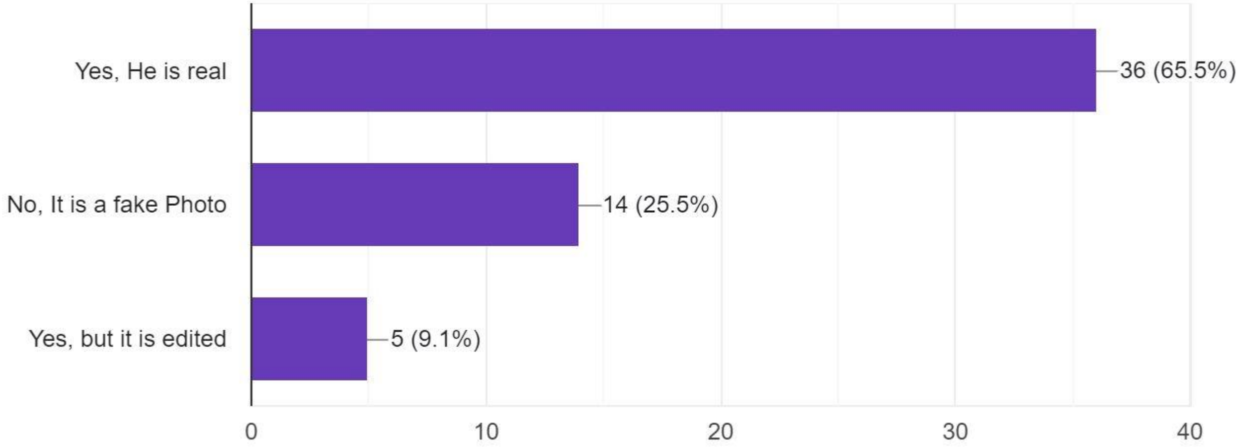
This Photo is fake and the person does not exist.

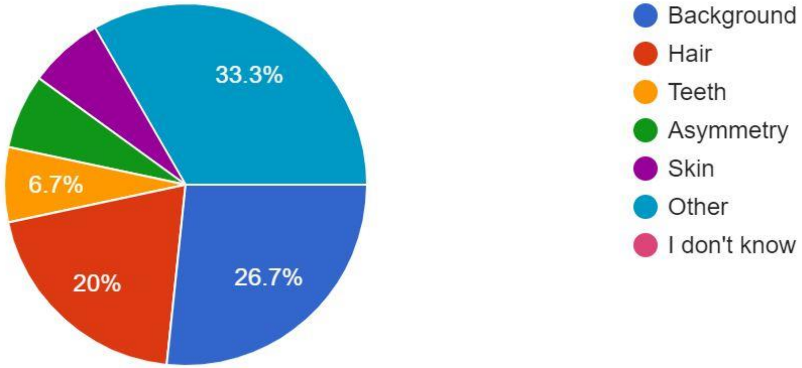




Real or *Fake* ?

This Photo is fake and the person does not exist.

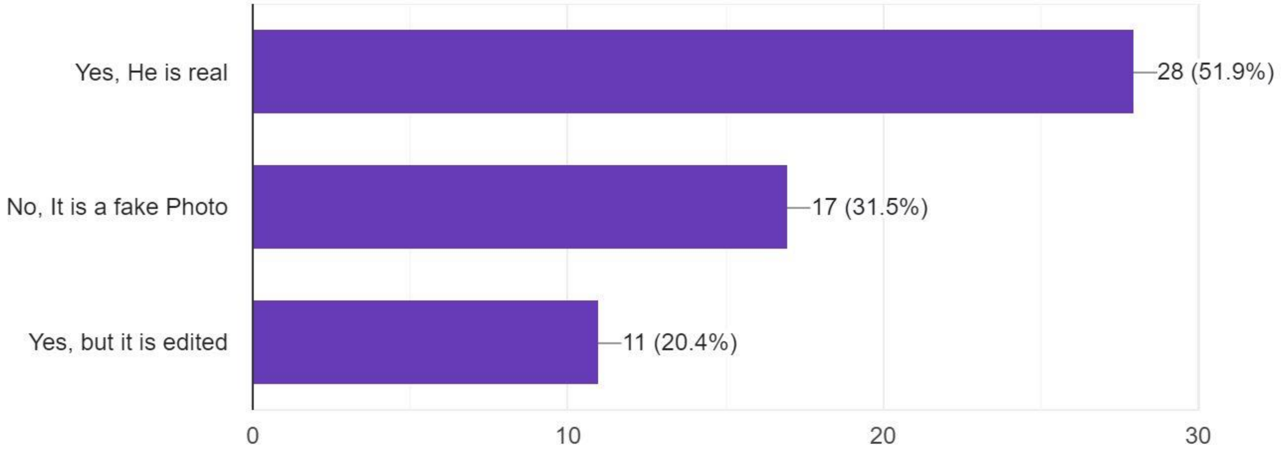


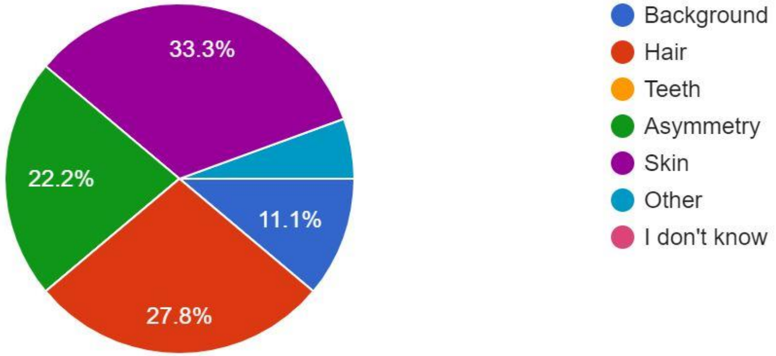


Zeinab Rahimi

Real or *Fake* ?

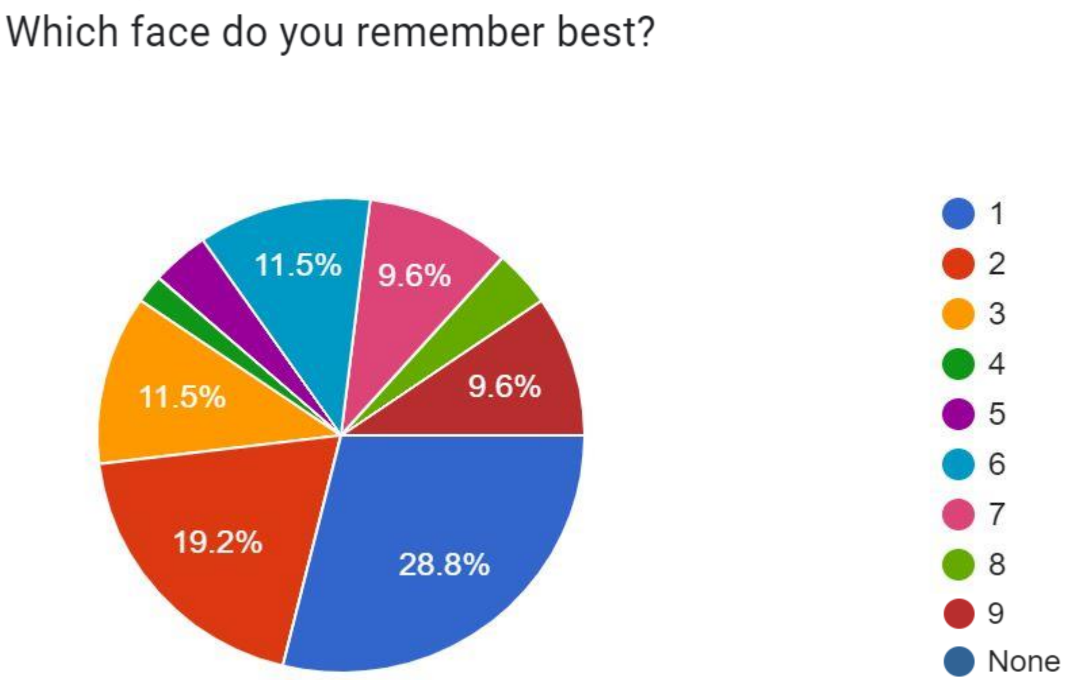
This Photo is fake and the person does not exist.





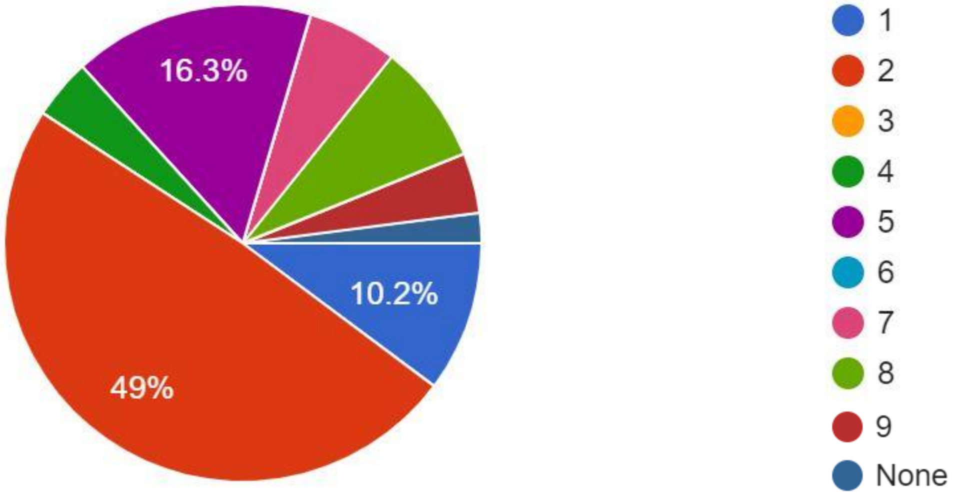
Real or *Fake* ?


1
2
3
4
5
6
7
8
9


Which face do you remember best?
11.5%
9.6%
9.6%
11.5%
19.2%
28.8%
1
2
3
4
5
6
7
8
9
None

Real or *Fake* ?

Which face do you trust more?



R e s u l t

Result :

First section of questionnaire:

A total of 55 participants participated in the survey. According to the results, the most incorrect answers are those related to the recognition of fake (computer-generated) photos, where the majority of respondents were not able to recognize them as being fake. When it comes to choosing a fake or edited picture, the skin is the most common reason cited.

Photo	Correct Answer	Correct answers from respondents (in percentage)	Wrong answers from respondents (in percentage)
1	Real but edited	60	40
2	Fake	18.2	81.8
3	Real	54.5	45.5
4	Real but edited	70.9	29.1
5	Fake	25.5	74.5
6	Real	85.2	14.8
7	Fake	31.5	72.3
8	Real	58.1	41.9
9	Real but edited	49.1	50.9

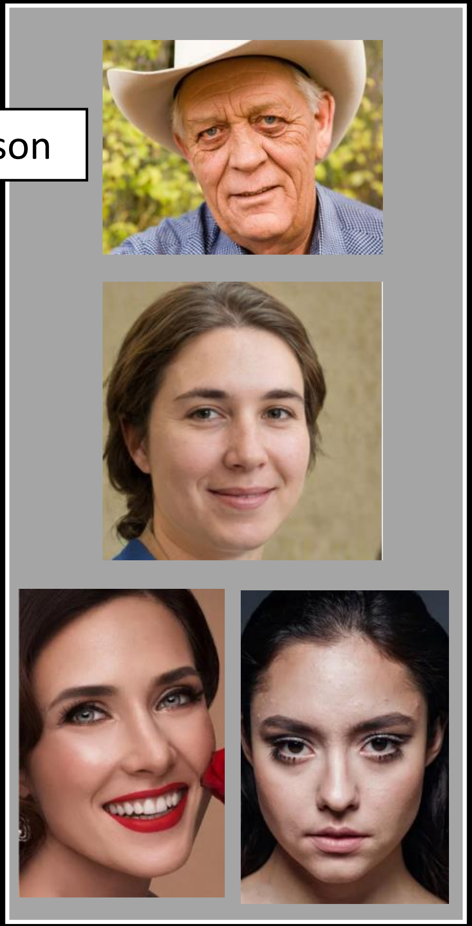
Table 1: Summary of the respondents' answers in percentage.

Result

Second section of questionnaire:

The result of the second section shows that the picture of an old man who is a real photo has the highest percentage of memorable pictures with 28.8%, but a fake photo is the second with 19.2%. The next two memorable pictures are edited photographs, each with the same percentage of 11.5%.

The most memorable Person



Real photo
28.8 %

Fake photo
19.2 %

Real but edited photo
11.5%

Real or *Fake* ?

Result

Third section of questionnaire:

In this section, participants gave a computer-generated photo the highest level of trust with 49%, while the second most trustable photo was also a fake photo with 16.3%, and the third photo, which is a real photo, received only a 10.2% trust rating.

Fake photo
49 %

The most trustable Person

Fake photo
16.3 %

Real photo
10.2%

Conclusion

Conclusion:

Based on the results, it can be concluded that respondents were most likely unable to recognize fake (computer-generated) photos but could recognize edited photos fairly well. One of the most common reasons for choosing a fake or edited picture is the skin, and respondents are relatively able to distinguish a non-real or altered skin from a real one. It could be considered a positive sign that the most memorable person was a real photo, according to the part two of the survey. In spite of this, most respondents chose fake photos over real ones when it comes to trust. This may indicate that people can develop positive feelings towards fake photos.

References :

1. https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/

2. T. T. Nguyen, Q. V. Hung Nguyen, D. Tien Nguyen, D. Thanh Nguyen, T. Huynh-The, S. Nahavandi, T. Tam Nguyen, Q. Pham, C. M. Nguyen, Deep learning for deepfakes creation and detection: A survey, Computer Vision and Image Understanding, 223, 2022(https://doi.org/10.1016/j.cviu.2022.103525.)

3. E. R. De Rezende, G. C. Ruppert and T. Carvalho, "Detecting computer generated images with deep convolutional neural networks", *2017 30th SIBGRAPI Conference on Graphics Patterns and Images (SIBGRAPI)*, pp. 71-78, 2017.

4. Bloomberg. How faking videos became easy and why that's so scary. https://fortune.com/2018/09/11/deepfakes-obama-video/, September 2018.

5. Robert Chesney and Danielle Citron. Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Affairs, 98:147, 2019.

6. T. Hwang. Deepfakes: A grounded threat assessment. Technical report, Centre for Security and Emerging Technologies, Georgetown University, 2020

7. . Xinyi Zhou and Reza Zafarani. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5):1–40, 2020.

8. Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. Deepfake: improving fake news detection using tensor decomposition-based deep neural network. The Journal of Supercomputing, 77(2):1015–1037, 2021.

9. Bin Guo, Yasan Ding, Lina Yao, Yunji Liang, and Zhiwen Yu. The future of false information detection on social media: New perspectives and trends. ACM Computing Surveys (CSUR), 53 (4):1–36, 2020.

10. Patrick Tucker. The newest AI-enabled weapon: 'deep-faking' photos of the earth. https://www.defenseone.com/ technology/2019/03/next-phase- ai- deep- fakingwhole-world-and-china-ahead/155944/, March 2019.

11. T Fish. Deep fakes: AI-manipulated media will be 'weaponised' to trick military. https://www.express. co . uk / news / science / 1109783 / deep - fakes - ai - artificial - intelligence - photos - video - weaponised-china, April 2019.

12. Marr. The best (and scariest) examples of AI-enabled deepfakes. https://www.forbes.com/sites/bernardmarr/ 2019/07/22/the-best-and-scariest-examples-ofai-enabled-deepfakes/, July 2019.

13. Yisroel Mirsky and Wenke Lee. The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR), 54(1): 1–41, 2021.

14. Luisa Verdoliva. Media forensics and deepfakes: an overview. IEEE Journal of Selected Topics in Signal Processing, 14(5): 910–932, 2020.