# Perspectives on DeepFakes for Privacy: Comparing Perceptions of Photo Owners and Obfuscated Individuals towards DeepFake Versus Traditional Privacy-Enhancing Obfuscation


Mohamed Khamis
mohamed.khamis@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Rebecca Panskus
rebecca.panskus@rub.de
Ruhr University Bochum
Bochum, Germany

Habiba Farzand
h.farzand.1@research.gla.ac.uk
University of Glasgow
Glasgow, United Kingdom

Marija Mumm
2232817m@student.gla.ac.uk
University of Glasgow
Glasgow, United Kingdom

Shaun Macdonald
shaun.macdonald@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Karola Marky
karola.marky@rub.de
Ruhr University Bochum
Bochum, Germany



**Figure 1: This paper investigates perceptions of photo owners and obfuscated individuals towards face obfuscation using blurring, pixelating, masking, avatar, and DeepFakes, i.e., using synthetically generated faces.**


## Abstract

Obfuscating people's faces using synthetically generated faces, i.e., DeepFakes, has been shown to be effective at privacy preservation. While recent work showed that DeepFake obfuscation is well perceived by viewers, the perspectives of a) the owner of the obfuscated photo, and b) the person that is being obfuscated, remain unclear. This paper reports on the results of a user study where participants uploaded their own group photos, in which they appear, and applied obfuscation techniques to both themselves and others in the image. The obfuscation methods included DeepFakes and four traditional techniques: blurring, pixelating, masking, and avatars. Our findings show that both photo owners and obfuscated individuals perceive DeepFake obfuscation as significantly more effective in protecting privacy compared to the traditional methods, and was found to integrate well with the environment.



*MUM '24, December 1–4, 2024, Stockholm, Sweden*




## CCS Concepts

• **Security and privacy → Economics of security and privacy**; **Privacy protections**; **Social aspects of security and privacy**;

## Keywords

Deepfakes, User Perceptions, Privacy, Anonymisation

**ACM Reference Format:**
Mohamed Khamis, Rebecca Panskus, Habiba Farzand, Marija Mumm, Shaun Macdonald, and Karola Marky. 2024. Perspectives on DeepFakes for Privacy: Comparing Perceptions of Photo Owners and Obfuscated Individuals towards DeepFake Versus Traditional Privacy-Enhancing Obfuscation. In *International Conference on Mobile and Ubiquitous Multimedia (MUM '24), December 1–4, 2024, Stockholm, Sweden.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3701571.3701602


## 1 Introduction

Taking and sharing photos online is part of everyday life today. When taking a photo in a public place, it is not uncommon that the photos include unintended subjects–known as bystanders. This raises privacy concerns, as bystanders may appear in these photos without their knowledge or consent. Some photo sharing portals have reacted to this, obfuscating the faces of individuals by blurring [32], pixelating [11, 29, 32], masking [29, 59], and replacing faces by avatars [40, 41] or cartoons [19].

More recently, AI-based approaches that use synthetically generated faces of individuals that do not exist were used for obfuscating individuals. We refer to this approach hereafter as *DeepFake obfuscation*. While there is controversy about DeepFakes due to its malicious use cases, DeepFake obfuscation can also be used to protect the privacy of individuals in shared photos while preserving the aesthetic quality of the image. Two aspects of DeepFake obfuscation have been studied. Khamis et al. [27] investigated the effectiveness of DeepFake obfuscation for privacy protection and demonstrated that it significantly outperforms other common obfuscation methods, such as blurring and pixelation, in concealing the identity of the obfuscated individual. This is complemented by the findings of Xu et al. [58] and Wöhler et al. [57], who investigated the effectiveness of DeepFakes in the context of content replacement and 360° videos. Their findings show that the majority of their participants were unable to detect the content replacement, and it being more effective compared to other obfuscation methods like blurring or masking [57, 58]. A second line of research has focused on perceptions of DeepFake obfuscation, showing that viewers perceive obfuscation with DeepFakes as the most realistic obfuscation method [57]. While previous studies [27, 57, 58] focused on the perceptions of the viewers, the perception of photo owners or the people that are obfuscated has not been investigated yet.

This shows that while previous work investigated the effectiveness of DeepFake obfuscation and its perception by viewers, it did not study a) the perceptions of the photo owner towards the obfuscated photos, and b) the perceptions of the individuals whose identities were obfuscated in those photos. This motivates our main research questions:

**RQ1:** How do photo owners perceive DeepFake obfuscation of bystanders in their photos in comparison to common obfuscation techniques?

**RQ2:** How do bystanders perceive DeepFake obfuscation of themselves in comparison to common obfuscation techniques?

To answer our research questions, we conducted a user study (N=18). First, we implemented a tool for obfuscating photos using blurring [32], pixelating [11, 29, 32], masking [29, 59], replacing faces by avatars [19, 40, 41] and DeepFakes.

In contrast to other studies, our participants provided group photos that include themselves instead of receiving prepared or staged photos. This ensures that participants can identify with the photo and its content in general and sets a realistic scenario. The photos contained our participants as the photo owner as well as bystanders. We then asked participants to first obfuscate bystanders and then themselves. Participants viewed and compared the different obfuscations and a baseline (no obfuscation).

Our results show that perceptions towards DeepFake obfuscation are similar to those of other obfuscation methods. All obfuscation methods have a negative impact on perceived aesthetics and information present in the photo. DeepFakes, however, result in obfuscations that better integrate into the photo than other methods.

Overall, the feedback from participants indicates that DeepFake obfuscation blends well with photos. However, it even blends too well that some were concerned about the ethical implications as it may mislead viewers. We conclude the paper by discussing the implications of privacy-aware DeepFake obfuscation and how to leverage this technology in an ethical way responsibly.

*Research Contribution:* In summary, we present the results of the first user study (N=18) to investigate the perceptions of photo owners and the obfuscated individuals towards DeepFake obfuscation in comparison to other established methods.

(1) **Investigation of DeepFake obfuscation:** We present the first in-depth study of privacy protection by DeepFake obfuscation considering different perspectives of photo owners and obfuscated bystanders.
(2) **Considerations of state-of-the-art obfuscation:** Our user study compares state-of-the-art obfuscation techniques to DeepFakes and a baseline (no obfuscation).
(3) **Replication package:** We further contribute our questionnaires and a prototype web platform for privacy-aware DeepFake obfuscation built using state of the art methods for synthetic face generation available at [blinded for submission] to allow other studies to use our implementation for replication and further investigations.

## 2 Related Work

This section provides a summary of related work detailing investigations of different obfuscation methods, research on choosing faces to obfuscate and privacy protection thought DeepFakes.

### 2.1 Photo Privacy on Social Media

Literature has shown that sharing photos on social media can have various privacy implications [1, 3, 23, 44]. Not only that photos can reveal sensitive information of the photo owner themselves [5], but also of bystanders or photo co-owners. Especially when multiple persons are in a photo, this can lead to multiparty privacy conflicts, based on different privacy preference of the different individuals [4, 46, 49]. Next to the deletion of shared photos, current coping strategies after regretting a shared photo [54] are co-owners untagging themselves [12] or photo owners changing the photos' privacy settings [1]. However, this strategies do not consider the privacy preferences of bystanders. Considering obfuscation methods before sharing a photo could prevent photo owners from running into the aforementioned problems.

### 2.2 Investigations of Obfuscation Methods

The first stream of research investigated the effectiveness of different obfuscation methods [17, 18, 27, 35] demonstrating that methods like blurring and pixelating are ineffective [35] as they might not completely conceal identities of individuals. Inpainting, which refers to substituting a person by the background, was found to be effective [35], yet it removes information that individuals were present in the photo. DeepFakes were also investigated and found to be effective in privacy protection from viewers perspective [58]. For obfuscating bystanders in 360° videos, face-swapping with synthetic faces was shown to be effective and best in preserving the videos realism compared to masking and blurring [57].

However, cues in the photo, e.g., a person's clothes, might impact effectiveness [27]. Different parts of people can be obfuscated. Studies showed that individuals perceive photos where only faces and not entire bodies are obfuscated to be more aesthetic [17]. Based

on this previous work [17, 18, 27, 35], we chose to compare the obfuscation methods blurring, pixelating, masking and adding an avatar to DeepFake obfuscation from the photo owners perspective.

Studies by Elagroudy et al. [14, 15] investigated the effect of privacy-aware obfuscations on user's memories showing that ambiguous life logs with obfuscations may distort memories, since photo sharing on social media can function as a mnemonic technique [20, 45, 52, 53]. This was attributed to the retrieval-induced forgetting phenomenon in which humans may recall incorrect details due to inaccuracies in the cues they are examining [9, 42].

A stream of research specifically investigated how faces for obfuscation are chosen because this depends on who took the photo [16, 17, 24, 33]. This research consequently developed methods to identify bystanders in photos using information from the person who took the photo [24], photo tags [33], or AI [16]. Li and Caine [34] proposed a system that automatically detects sensitive content and obfuscates it by masking, blurring, an avatar or inpainting. Alharbi et al. [2] investigated the impact of obfuscation in general on visually impaired users showing that participants disliked automatic obfuscation without user interaction. These results partly replicated the research on choosing individuals to obfuscate detailed above [16, 17, 24, 33].

## 2.3 Privacy Protection through DeepFakes

When looking at research on privacy protection through DeepFakes in particular, there are four different categories of manipulating photos: (1) attribute manipulation, (2) expression swap, (3) identity swap, and (4) entire face synthesis [50].

Attribute manipulation uses Generative Adverserial Networks (GANs) to edit faces by changing specific features like age, hair, or skin colour [50]. However, when the extent of changes by this technique is too low, privacy cannot be ensured. Same applies to expression swapping, where the facial expression of the person is manipulated, but the persons identity remains preserved [50]. Identity swapping goes a step further as it not only changes the person's facial expression but also the whole face by replacing it by the face of a different person [50]. While for this technique a second person's face is needed, applying an entire face synthesis uses a non-existing face [50]. Hukkelas et al. [22] developed the *DeepPrivacy* face anonymisation architecture using a GAN for this purpose. With the technique of entire face synthesis compared to the other three, the highest level of obfuscation is reached. Face synthesis combined with photo access rights based on social graphs can be a first step towards automated privacy enhancing DeepFake techniques on social media platforms [8].

> **Summary:** In summary, related work investigated the effectiveness of state-of-the-art obfuscation methods showing that obfuscation also impacts the memory of individuals. We add an investigation of user perceptions about DeepFake obfuscation in comparison to the state-of-the-art obfuscation methods detailed above.

# 3 Method

The aim of our study was to gauge photo owners' perception of DeepFake obfuscation in comparison to other obfuscating techniques. The perception was analysed from two perspectives: first, when DeepFake obfuscation is applied to other individuals' faces and second when DeepFake obfuscation is applied to the participant's own face.

## 3.1 Captured Data

The study was designed as a within-subjects experiment with one independent variable: the OBFUSCATION METHOD. There were six conditions: 1) original photo (baseline), 2) blurring, 3) pixelating, 4) masking, 5) DeepFake and 6) avatar. The order of conditions was counterbalanced using a Latin Square. The study was conducted online via Zoom due to COVID-19 restrictions and abided by our university's ethics requirements. There were six experimental conditions (see also Figure 1):

- **BASELINE:** The original photo without any obfuscation applied. This served as the baseline.
- **DEEPFAKES:** We implemented DeepFake obfuscation using the DeepPrivacy framework by Hukkelås et al. [21, 22]. Their algorithm uses a GAN to generate fake faces while incorporating "style transfer" which allows customising a fake face by imposing some facial characteristics, such as the skin and hair colour of another person. In our use case, this means that the generated fake face has the same hair and skin colour as the obfuscated person, but still looks different. The DeepPrivacy framework considers the background and the pose of the face to create
- **BLURRING:** Li et al. [35] used a Gaussian blur with a radius of 4 pixels. To replicate these conditions, we used the *GaussianBlur()* method of the ImageFilter Module [36], setting the blurring radius to be directly proportional to the photo's width × height, with photos of size 770 × 552 pixels having a blurring radius of 4.
- **PIXELATING:** Pixelating was applied by downscaling the face to 15 × 15 pixels [35] and then scaling it up again to its original size but in a pixelated form. For smaller photos, we set the size parameter to be directly proportional to the product of the photo's width and height.
- **MASKING:** We applied a black rectangle on the individual's face. [35].
- **AVATAR:** We used emojis [28] instead of a human avatar as it is neutral to gender and skin colour. The emoji was resized to the size of the located face and placed over the face area.

We further captured the following dependent variables by having the participants rate the statements on a 5-point Likert-scale (1: "strongly disagree"; 5: "strongly agree"). The statements were based on prior work on privacy-aware obfuscation [10, 18, 35, 38, 43] and were modified to fit our study. The captured variables are:

- **LIKEABILITY:** We investigated the participants' likeability of the obfuscation methods similar to related work that investigated non-DeepFake methods [35, 35, 38] using the statement: "*I like this obfuscation technique.*"
- **AESTHETICS:** Since obfuscation impacts the appearance of the individuals in the photo and the overall composition, we asked: "*This photo is aesthetically pleasing.*", motivated by the related studies [10, 18, 35, 35].
- **PRIVACY PROTECTION:** Since the primary purpose of obfuscation is privacy protection, we asked participants to rate:

"*My privacy is protected.*", motivated by an effectiveness study of obfuscation methods [35].

**Comfort:** Related work also showed that obfuscation can impact a person's comfort when looking at (obfuscated) photos [35] resulting in the following statement in our study "*I am comfortable appearing like that in someone's photo available online.*"

**Integration:** Primary worked showed that people consider the overall composition of the photo particularly focusing on the integration of the obfuscation method [35]: "*My obfuscated face's features are well integrated within the photo.*"

**Information Availability:** Finally, obfuscation methods could conceal further information in photos [18, 35, 43] which is why we asked participants to rate: "*This obfuscation hides important information from the photo.*"

We further asked the participants some open-ended questions: (1) which of the obfuscation techniques they would like others to use on their faces when posting online and (2) to rank the obfuscations from the most to the least preferred. Finally, we asked participants for their opinions about using DeepFakes for privacy protection.

## 3.2 Apparatus

This section briefly details the implementation of our Photo Obfuscator App that we used in our study to obfuscate the photos. For comparability with the previous work [35], we contacted the authors to use the same parameters they used in their implementations of blurring, masking, and pixelating obfuscations.

Our web app is realised with the Python web framework Django and hosted on the PythonAnywhere service[1]. The landing page invited the user to upload a photo and apply different obfuscation techniques. After uploading, the photo is processed to locate face coordinates. For the purpose of our study, the form on the landing page is pre-filled with a participant ID generated by calculating the Unix timestamp in milliseconds.

Next, the user is presented with the original photo, displaying the identified faces (see Figure 2). The user then selects the face(s) they wish to obfuscate. The next page displays obfuscated versions using the techniques detailed above. To replicate the conditions from the study by Li et al. [35] who used $770 \times 552$-pixel photos, we resized the smaller side of the photo to 552 pixels and the larger side was resized to maintain the original aspect ratio.

## 3.3 Procedure

The study was conducted online and its procedure was as follows.

*Step 1: Consent and Previous Experience.* After expressing their consent, participants provided their demographics, shared their photo uploading habits, and whether they were involved in requests to remove photos from social media. Next, the participants were asked to navigate to the Photo Obfuscator App website to complete the following steps.

*Step 2: Obfuscating Others.* First, to investigate the photo owner's perception of the obfuscation methods, we simulated a situation in which the photo owner posts a photo containing other individuals online. Participants were asked to upload a photo they owned with at least two individuals. This was a recent group picture or selfie with others provided by the participant. Figure 2 shows an illustration of such a picture. After face detection, the participant was asked to select at least one person other than themselves to obfuscate. The participants then viewed six photos that represent the six conditions of this study. Next, participants filled in the questionnaire detailed above.

*Step 3: Obfuscating the Participant.* This step investigated the perceptions of appearing obfuscated in someone else's photo. The participants were asked to upload a different photo that contains themselves and at least one other person. This time the participants were asked to select their own face for obfuscation, and proceeded the same way as in Step 2.

*Step 4: Closing.* We concluded by collecting qualitative feedback, participants' ranking of the methods, and participants' opinions about using DeepFakes for privacy protection.

## 3.4 Participants

We recruited 18 participants aged between 19 and 28 years (Mean=23.33, SD=2.29) through mailing lists, social media, word-of-mouths and flyers. The recruitment materials did not mention DeepFakes to avoid biasing our participants.

Half of the participants identified as male, the other half as female. We provided further answer options, yet no participant chose them. All participants had social media accounts. The majority posted photos or disappearing photos (stories) of themselves or others every few months (N=7), followed by monthly (N=3), yearly (N=3), every few years (N=3), more than once a day (N=1) and weekly (N=1). Six participants reported to have contacted someone with a request to remove a photo from social media in which they appeared. Four participants reported having been contacted by someone else with a similar request.

## 3.5 Ethical Considerations

The study was conducted online and abided by our university's ethics requirements and was approved by our IRB. Since participants uploaded their private photos during the study, the photos were not stored on our server. Once the participant closed the website, the photo was no longer available to us. Before the study, participants were informed about their rights as study participants and given the opportunity to ask questions to the experimenter.

## 3.6 Limitations

In this section, we reflect on the limitations of our study. Our study relied on self-assessments of participants which might be susceptible to common biases, e.g., social acceptability. While the participants obfuscated their own photos, they might react differently if they saw the same photo uploaded by someone else on a social network.

Further, we used specific implementations used by related work to create comparable results. Yet, the values for blurring and pixelating were dependent on the dimensions of the photo. The face sizes

[1]Link redacted for anonymity. We will add it to the Camera Ready version.

**Figure 2: The participants uploaded photos and selected which faces they would like to obfuscate.**

on the uploaded photos varied and faces that were located closer to the camera might have been more recognisable after applying the aforementioned obfuscation methods compared to the faces located further away. Another limitation is the use of a single smiling emoji for avatar obfuscation. Perceptions could have been different had we used a set of different emojis.

## 4 Results

We first checked assumptions for statistical analysis. Since our data was not normally distributed, we analysed the Likert items using Friedman tests. For the post-hoc analysis, we applied pairwise comparisons using Wilcoxon tests. Bonferroni correction was applied due to multiple comparisons resulting in a significance level set at $p < 0.00333$. We report means and standard deviations in the text, while medians can be interpreted from the Likert plots depicted in Figure 3. For the detailed statistics, the reader is referred to Appendix A.

### 4.1 Perceived Likeability

In this section, we report the results of how the participants liked the different obfuscation methods. For both, obfuscating others and obfuscating oneself, we could not find statistically significant effects of the OBFUSCATION METHOD on its likeability ($p > 0.05$ each).

*4.1.1 Obfuscating Others.* When comparing the mean score for each method, we find that PIXELATING received the highest score (M=3.72, SD=0.93), followed by DEEPFAKES (M=3.28, SD=1.52), BLURRING (M=2.94, SD=1.18), AVATAR (M=2.89, SD=1.56), MASKING (M=2.67, SD=1.25). The participants were also asked to rate how they like the BASELINE (M=2.11, SD=1.37) considering that this does not offer any obfuscation. The BASELINE scored lowest.

*4.1.2 Obfuscating Self.* DEEPFAKES had the highest scores (M=3.5, SD=1.38), followed by PIXELATING (M=3.17, SD=1.21), BLURRING (M=2.94, SD=1.22), MASKING (M=2.83, SD=1.54), AVATAR (M=2.61, SD=1.3). The participants were again asked to rate how they like the BASELINE which again scored lowest (M=2.5, SD=1.57).

### 4.2 Perceived Aesthetics

Considering aesthetics, we investigated how aesthetically pleasing participants considered the obfuscations.

*4.2.1 Obfuscating others.* A Friedman test revealed a statistically significant effect ($\chi^2(5) = 33.927, p < 0.001$). Significant differences were observed between the BASELINE and each OBFUSCATION METHOD: BLURRING ($Z = -3.552, p < 0.001$), PIXELATING ($Z = -3.337, p = 0.001$), MASKING ($Z = -3.769, p < 0.001$), DEEPFAKES ($Z = -3.117, p = 0.002$), and AVATAR ($Z = -3.335, p = 0.001$). Participants found the BASELINE photos to be the most aesthetically pleasing (M=4.72, SD=0.56). Lower scores were given to PIXELATING (M=3.11, SD=1.1), DEEPFAKES (M=2.78, SD=1.51), BLURRING (M=2.78, SD=1.27), AVATAR (M=2.5, SD=1.21), and MASKING (M=2.11, SD=0.9).

*4.2.2 Obfuscating Self.* We again found a statistically significant effect ($\chi^2(5) = 41.958, p < 0.001$). Significant differences were observed between the BASELINE and each of the OBFUSCATION METHODS: BLURRING ($Z = -3.575, p < 0.001$), PIXELATING ($Z = -3.579, p < 0.001$), MASKING ($Z = -3.677, p < 0.001$), DEEPFAKES ($Z = -3.028, p = 0.002$), and AVATAR ($Z = -3.559, p < 0.001$). The BASELINE scored the highest (M=4.78, SD=0.42), followed by DEEPFAKES (M=3.28, SD=1.41), PIXELATING (M=3.0, SD=1.0), BLURRING (M=2.94, SD=1.08), AVATAR (M=2.28, SD=1.19), and MASKING (M=2.17, SD=1.12).

### 4.3 Perceived Privacy Protection

Unsurprisingly, the BASELINE was perceived to be the least privacy-preserving when obfuscating oneself and others. All differences between the BASELINE and all other OBFUSCATION METHODS were significant. Below, we report the detailed results.

*4.3.1 Obfuscating Others.* The Friedman test revealed a statistically significant effect ($\chi^2(5) = 52.628, p < 0.001$). Significant differences were observed between the BASELINE and each of the OBFUSCATION METHODS: BLURRING ($Z = -3.271, p = 0.001$), PIXELATING ($Z = -3.557, p < 0.001$), MASKING ($Z = -3.787, p < 0.001$), DEEPFAKES ($Z = -3.703, p < 0.001$), and AVATAR ($Z = -3.674, p < 0.001$). Significant differences were also found between MASKING and BLURRING ($Z = -3.108, p = 0.002$).

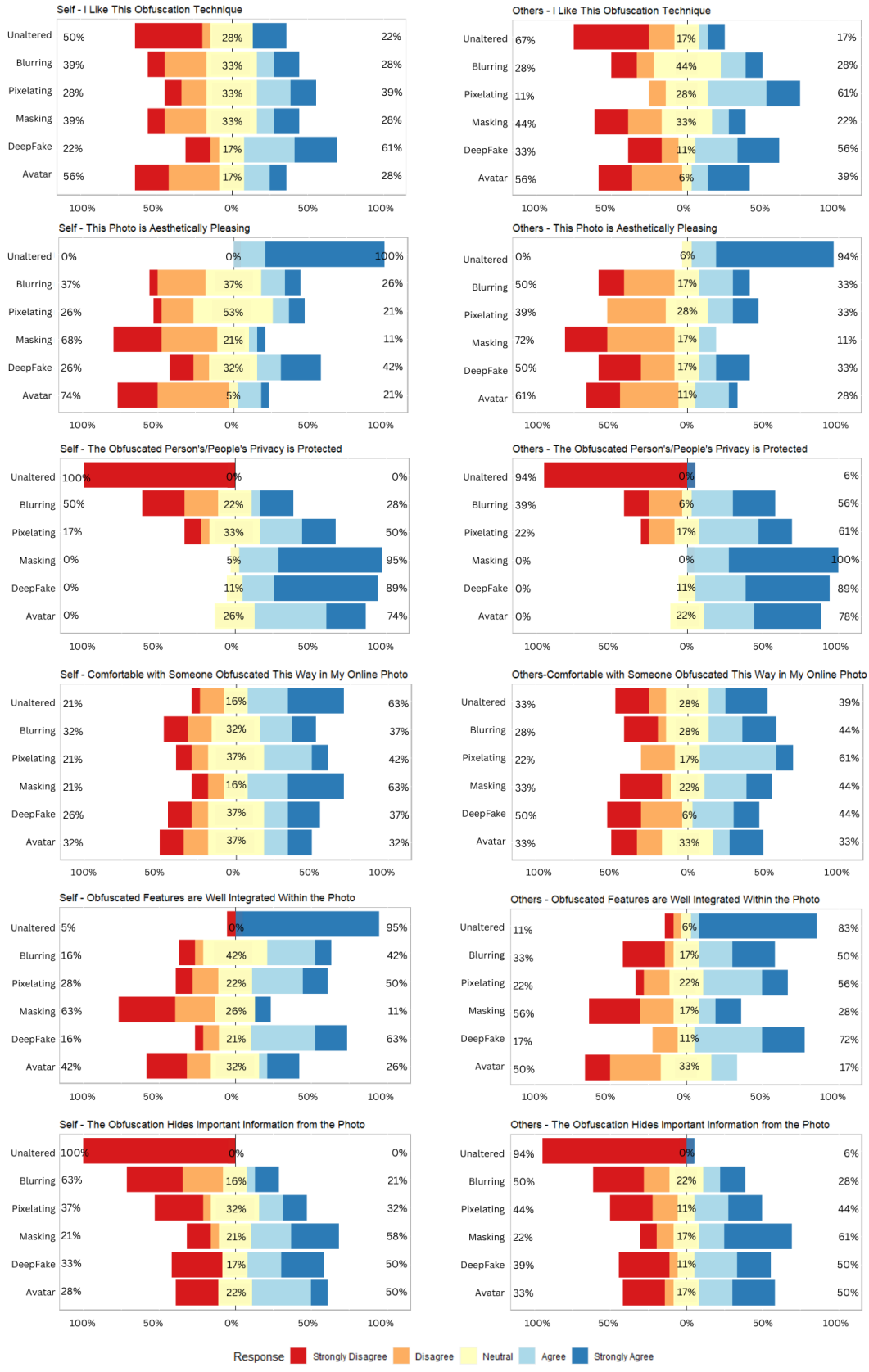


**Figure 3: Results of the Likert items. The plots on the left side, consider obfuscating oneself. The one on the right consider obfuscating others.**

Participants' perceived privacy of the BASELINE lowest (M=1.22, SD=0.92), and highest when using MASKING (M=4.72, SD=0.45), followed by DEEPFAKES (M=4.44, SD=0.68), AVATAR (M=4.22, SD=0.79), PIXELATING (M=3.56, SD=1.17), and BLURRING (M=3.28, SD=1.48).

*4.3.2 Obfuscating Self.* We found a statistically significant effect of the OBFUSCATION METHOD on perceived privacy in photos where others are obfuscated $\chi^2(5) = 60.253, p < 0.001$. Significant differences were observed between the BASELINE and each of the obfuscated versions: BLURRING ($Z = -3.209, p = 0.001$), PIXELATING ($Z = -3.555, p < 0.001$), MASKING ($Z = -3.866, p < 0.001$), DEEPFAKES ($Z = -3.862, p < 0.001$), and AVATAR ($Z = -3.780, p < 0.001$). Significant differences were also found between MASKING and each of BLURRING ($Z = -3.204, p = 0.001$) and PIXELATING ($Z = -3.114, p = 0.0032$), and also between DEEPFAKES and each of BLURRING ($Z = -3.050, p = 0.002$) and PIXELATING ($Z = -2.939, p = 0.003$).

Similar to when obfuscating others, MASKING had the highest score (M=4.61, SD=0.59), followed by DEEPFAKES (M=4.56, SD=0.68), AVATAR (M=4.0, SD=0.75), PIXELATING (M=3.44, SD=1.21), BLURRING (M=2.72, SD=1.48), and THE BASELINE (M=1.0, SD=0.0).

## 4.4 Perceived Comfort

We found no statistically significant effect on the participants' COMFORT with someone or the self appearing obfuscated in photos ($p > 0.05$, each). This indicates that the participants do not consider different levels of comfort when being obfuscated by different methods.

*4.4.1 Obfuscating Others.* Participants scored PIXELATING (M=3.5, SD=0.96) the highest, followed by BLURRING (M=3.17, SD=1.42), the BASELINE (M=3.11, SD=1.49), AVATAR (M=3.06, SD=1.35), MASKING (M=3.0, SD=1.45), and DEEPFAKES (M=2.89, SD=1.45).

*4.4.2 Obfuscating Self.* Participants were most comfortable with the BASELINE photo of themselves (M=3.72, SD=1.28), followed by MASKING (M=3.67, SD=1.37), PIXELATING (M=3.22, SD=1.13), DEEPFAKES (M=3.17, SD=1.34), BLURRING (M=3.06, SD=1.31), and AVATAR (M=3.0, SD=1.29).

## 4.5 Perceived Integration of Obfuscation within the Photo

*4.5.1 Obfuscating Others.* Statistical analysis revealed a statistically significant effect ($\chi^2(5) = 23.369, p < 0.001$). Significant differences were observed between the BASELINE and AVATAR ($Z = -3.033, p = 0.002$), and between DEEPFAKES and AVATAR ($Z = -3.031, p = 0.002$). Participants reported they found the highest integration in the BASELINE (M=4.44, SD=1.17), followed by DEEPFAKES (M=3.83, SD=1.01), PIXELATING (M=3.44, SD=1.12), BLURRING (M=2.89, SD=1.24), MASKING (M=2.56, SD=1.46), AVATAR (M=2.5, SD=0.96).

*4.5.2 Obfuscating Self.* Similar to the above, we found a statistically significant effect of the obfuscation method on the perception of how well the obfuscated person(s) are integrated within the photo $\chi^2(5) = 33.393, p < 0.001$. Significant differences were observed between the BASELINE and all OBFUSCATION METHODS except DEEPFAKES (BLURRING ($Z = -3.046, p = 0.002$), PIXELATING ($Z = -3.336, p = 0.001$), MASKING ($Z = -3.410, p = 0.001$), and AVATAR, ($Z = -3.078, p = 0.002$)).

The participants ranked the techniques in almost the same order as above. The BASELINE was ranked the highest (M=4.78, SD=0.92), followed by DEEPFAKES (M=3.61, SD=1.11), PIXELATING (M=3.28, SD=1.24), BLURRING (M=3.28, SD=1.1), AVATAR (M=2.78, SD=1.47), MASKING (M=2.22, SD=1.27).

## 4.6 Photo Information Sufficiency

*4.6.1 Obfuscating Others.* We found a statistically significant effect of the OBFUSCATION METHOD on the perception of hiding important information ($\chi^2(5) = 36.410, p < 0.001$). Significant differences were found between the BASELINE and four OBFUSCATION METHODS: PIXELATING ($Z = -3.082, p = 0.002$), MASKING ($Z = -3.453, p = 0.001$), DEEPFAKES ($Z = -2.971, p = 0.003$) and AVATAR ($Z = -3.088, p = 0.002$).

Participants rated that the BASELINE to hide the least information (M=1.22, SD=0.92), followed by BLURRING (M=2.61, SD=1.46), PIXELATING (M=2.94, SD=1.54), DEEPFAKES (M=3.0, SD=1.6), AVATAR (M=3.17, SD=1.57), and MASKING (M=3.72, SD=1.41).

*4.6.2 Obfuscating Self.* Finally, we also found a statistically significant effect of the OBFUSCATION METHOD on the perception of hiding important information ($\chi^2(5) = 43.036, p < 0.001$). Significant differences were found between the BASELINE and four OBFUSCATION METHODS: BLURRING ($Z = -2.961, p = 0.003$), PIXELATING ($Z = -3.093, p = 0.002$), MASKING ($Z = -3.446, p = 0.001$), DEEPFAKES ($Z = -3.100, p = 0.002$) and AVATAR (Z = $-3.247, p = 0.001$). Results were relatively similar to obfuscating others, with AVATAR and DEEPFAKES switching places in the ranking: BASELINE (M=1.0, SD=0.0), BLURRING (M=2.39, SD=1.46), PIXELATING (M=2.78, SD=1.47), AVATAR (M=3.06, SD=1.39), DEEPFAKES (M=3.11, SD=1.63), MASKING (M=3.5, SD=1.42).

> **Summary:** We could not find large differences in perception when obfuscating others or oneself. We could not find evidence of impact on LIKEABILITY, or COMFORT in general. In terms of INTEGRATION with photo features, DEEPFAKES were perceived significantly better than AVATAR. Further, DeepFake obfuscation is perceived to PROTECT PRIVACY significantly better than the BASELINE and was the only method that we did not find evidence of a significant reduction in INFORMATION AVAILABILITY compared to the BASELINE. Nevertheless, DEEPFAKES significantly worsen perceived photo AESTHETICS.

## 4.7 Qualitative Feedback

Participants provided feedback in a free-text field at the end of the study. We analyzed the collected feedback by inductive qualitative content analysis according to Kuckartz [30]. To ensure reliability, the formation and assignment of categories were discussed with two other researchers and final code allocations were agreed on.

All participants commented on all techniques but provided more feedback about DeepFakes, likely due to their novelty.

*4.7.1 Opinions on Avatars are Divided.* Regarding AVATAR obfuscation, some participants believed using the emoji avatar is the most appropriate for social media and the only technique that would not

make people feel uneasy. Others considered its use "*inappropriate*" since it might depict an emotion that does not match that of the individual.

*4.7.2 Experiences are Important.* Some participants mentioned associating PIXELATING with criminal activity, or that PIXELATING is *"usually associated with pornography"* (P103). We conclude, that the context of use of an OBFUSCATION METHOD impacts the perceptions of participants. While DEEPFAKES would not be used to obfuscate criminals, other well-known techniques might be.

*4.7.3 DeepFakes are Fun yet Challenging to Spot.* Participants gave positive feedback about DEEPFAKES, stating that it has a *"natural-looking result"* (P201) and it was the *"only acceptable obfuscation technique"* (P401). Others commented that they *"laughed too hard"* (P302) or found it *"fun"* (P601, P202, P404) to use or *"interesting"* (P303, P404) especially when applied on themselves P404: *"If I wanted my face to be obfuscated, deepfake seems like a lot of fun"*.

One participant raised a point that *"DeepFake technique makes it least possible to guess that some obfuscation technique has been applied"* (P602), suggesting that the obfuscation is well integrated with the photo. Indeed, participants expressed difficulties in determining the difference between a real photo and a manipulated one. For example, P102 stated: *"Deepfake, while masking the person's identity seems unpleasant and others may mistake it for the person's real face"*. A similar statement was given by P103: *"using DeepFake is like pretending [to be] someone else"*. On the downside, P302 remarked that DEEPFAKES, if not marked, might lead to misinterpretation: *"could cause misinterpretations e.g., I am in a photo with someone who is not actually there"*. Some participants found that DEEPFAKES *"looked scary"* (P503), and expressed concern that *"generated faces are not very pleasing"* (P402). One participant was concerned DEEPFAKES might cause negative feelings *"DeepFake obfuscation might be offensive to my friend(s), as they may assume they do not look good enough to appear in my photos posted on social media"* (P504).

Finally, participants regarded DEEPFAKE obfuscation as a method for protecting privacy online in an unintrusive way: *"... without making it obvious that you wanted to remove someone's face from the photo to protect [them]"* (P101). Hence, it was considered to be appropriate for *"public online spaces such as maps, reviews and promotions"* (P401). One of the participants said that they would only support the idea if their face in the photo was not swapped for a real person's face. This is already the case in our implementation as we replace the individual's face with a synthetically generated one.

# 5 Discussion

In this section, we discuss how photo owners perceive DeepFakes as a privacy protector.

## 5.1 RQ1 and RQ2: Perceptions of DeepFake Obfuscation?

In our study, we investigated how (a) photo owners perceive the obfuscation of bystanders in their photos and (b) how bystanders perceive the obfuscation of themselves in others' photos. In a nutshell, our results indicate that users' perceptions towards obfuscating themselves compared to obfuscating others are largely similar. We could not find evidence that users like or feel more comfortable using one of the obfuscation methods we compared over the others. However, unsurprisingly, all obfuscation methods scored worse than the baseline in terms of perceived aesthetics and information sufficiency. Privacy protection is perceived as significantly higher when using DeepFakes or masking, which is also in line with the results from related effectiveness studies [27, 35]. Avatars were perceived less positively than DeepFakes and the baseline in terms of integration with the rest of the photo. This might be rooted in our implementation of avatars because we used emojis based on prior work [35, 40, 41].

Feedback from participants indicates that DeepFakes are effective and covert. Several participants reported it is difficult to determine that a photo has been tampered with because the DeepFakes integrated with the rest of the photo. As a consequence, participants had concerns about how this could mislead viewers. The quality of DeepFakes was an important metric for participants who reported they may not use it if the quality is low or uncanny.

Our results show that perceptions of photo owners and obfuscated individuals are positive, which is in line with prior findings on the perceptions of the viewers of photos towards similar implementations of DeepFake obfuscation [57, 58].

## 5.2 Should we Obfuscate with DeepFakes?

There have been several investigations into obfuscation with DeepFakes in the past (cf. [34, 58]). While the privacy improvements are obvious [58], the question arises whether and how DeepFake algorithms should be used for obfuscation. This section discusses implications of using DeepFakes and provides guidance for future work.

*5.2.1 The Quality of the Fake Faces Impacts Perceptions in Both Ways.* The quality of the fake faces generated by the DeepFake frameworks, such as the DeepPrivacy framework [22], can vary. This might result in a face that seems "strange", unnatural, or cause the "uncanny valley effect", i.e. the feeling of revulsion due to seeing an imperfect resemblance of a human [37]. This effect may be intensified due to the "perceptual mismatch" effect, which happens when there are discrepancies between expected and actual visual cues, particularly in faces that appear almost, but not entirely, human [31]. In the context of DeepFake obfuscation, this effect could occur when individuals view their own photos featuring familiar faces, or when they recognise themselves in a photo where their face has been obfuscated as also shown by participant comments in our study.

As methods for face synthesis improve, the quality of DeepFakes will. However, there might still be cases where algorithms do not perform well due to, for example, biased training sets. While we did not encounter such issues during our study, we did encounter issues in our pilot testing where, for example, an adult's face was replaced by a child's. Based on that, we formulate our first takeaway:

> **Takeaway 1:** *Systems that employ DeepFakes for obfuscation should allow photo owners to generate a new face if they are not satisfied with the result.*

*5.2.2 Controversies Surrounding DeepFakes.* One of the topics brought up in our participants' feedback was how the obfuscation techniques are used in other domains. Pixelating, in particular, reminded some participants of convicts as their faces are usually obfuscated this way when they appear in the media. One participant even mentioned that DeepFakes are associated with pornography. Because of that, the public might stigmatise DeepFakes because of their potential misuses, such as fake news and involuntary pornography. This is in line with research in law and policy, which emphasised how the misuse of DeepFakes in inappropriate contexts can create unease and mistrust, especially when individuals find themselves involved in content manipulated by this technology [7]. In the context of DeepFake obfuscation, photo owners may be hesitant to use that technology on their acquaintances who may appear in the photos, fearing that it could invoke negative associations or discomfort among those who appear in the photos. Similarly, the obfuscated individuals may also experience a sense of unease or mistrust towards the photo owners when they see themselves represented through a controversial technique like DeepFakes due to its association with inappropriate content.

However, the perception towards DeepFakes in general might shift when the public learns more about their useful applications. DeepFakes have several benefits in various sectors, such as entertainment, education, healthcare, and digital communications [55]. For example, DeepFakes have been used to change the visual appearance of celebrities e.g., to de-age them, and to digitally resurrect deceased celebrities and family members [6, 13, 39, 56].

> **Takeaway 2:** *Experiences with obfuscation techniques and their reputation play an integral role in their perception.*

*5.2.3 Marking DeepFakes and Ethical Implications.* While DeepFakes are a promising way to protect privacy, they may also be leveraged for unethical use, such as impersonation [48]. For this, it is essential to only use *synthetic* faces instead of faces from real individuals. Yet, no matter what kind of face is used, it might be challenging or even impossible for users to spot DeepFakes [58]. When DeepFake websites[2] were launched, discussions on whether the synthetically generated faces may coincidentally look like a real person started as well. There have been no formal studies of the likelihood of this to happen, likely due to the infeasibility of conducting such a study. If this does occur too often, then this would be a limitation of DeepFakes obfuscation.

Before DeepFake obfuscation becomes widely available to the public, we argue that there is a need to "mark" that photos have been tampered with whenever any obfuscation method is used. One particularly promising way to achieve this is by declaring this in the metadata of the generated file [25, 26] allowing systems to automatically scan this information and warn viewers. Other suggestions are preserving provenance and attribution data for digital content to counter misinformation [25] or using Blockchain technology to facilitate tracking the origin of photos and the changes that have been made to them [51]. On the non-technical side, contextualised training and education have also been shown to assist in raising awareness and detection of DeepFakes [47].

> **Takeaway 3:** *Photos with DeepFake obfuscation should use synthetic faces and should be clearly marked to indicate that they have been manipulated.*

Based in the discussion above and our takeaways, we conclude that DeepFakes are a promising solution for effective obfuscation that keeps the aesthetics of a photo. However, future work should investigate the likelihood that a synthetic face matches the one of a real person and effective means to mark and detect manipulated photos.

## 6 Conclusion

This paper presented a user evaluation of DeepFake obfuscation compared to common obfuscation methods. Our results show that DeepFake obfuscation is promising in terms of privacy protection and photo aesthetics. Perceptions towards DeepFakes are largely similar to those towards other methods. Yet, DeepFakes are better integrated with the rest of the photo. However, DeepFakes are a technology that can be used in different directions. While our work highlights the potential for aesthetic photos that protect the privacy of bystanders in social media, we need further investigations on the potential misuses of DeepFakes and their ramifications. Future work should investigate methods for marking DeepFakes, communicating DeepFaked content to users, and methods to detect DeepFakes in non-marked photos.

## Acknowledgments

This work was co-funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2092 CASA - 390781972, the joint project: digital fitness for citizens - realistic risk perception, secure routines - "DigiFit" - grant number 16KIS1646K, and the program "Digitalisierungsprofessuren für Niedersachsen", an EPSRC New Investigator Award (grant number EP/V008870/1), by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has also been funded by the UK EPSRC under grant number EP/S035362/1, and an Excellence Bursary Award by the University of Glasgow.

## References

[1] Shane Ahern, Dean Eckles, Nathaniel S. Good, Simon King, Mor Naaman, and Rahul Nair. 2007. Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 357–366. https://doi.org/10.1145/1240624.1240683

[2] Rahaf Alharbi, Robin N. Brewer, and Sarita Schoenebeck. 2022. Understanding Emerging Obfuscation Technologies in Visual Description Services for Blind and Low Vision People. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 469 (nov 2022), 33 pages. https://doi.org/10.1145/3555570

[3] Andrew Besmer and Heather Richter Lipford. 2010. Moving beyond untagging: photo privacy in a tagged world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1563–1572.

[4] Gergely Biczók and Pern Hui Chia. 2013. Interdependent privacy: Let me share your data. In *Financial Cryptography and Data Security: 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers 17*. Springer, 338–353.

[5] Livio Bioglio and Ruggero G Pensa. 2022. Analysis and classification of privacy-sensitive content in social media posts. *EPJ Data Science* 11, 1 (2022), 12.

[2] E.g., https://www.thispersondoesnotexist.com/ last accessed Aug-21-2024

[6] Campaign. 2019. What are the laws against deepfake pornography across the UK? https://www.campaignlive.co.uk/article/deepfake-voice-tech-used-good-david-beckham-malaria-campaign/1581378 Retrieved November 3, 2021.

[7] Robert Chesney and Danielle Citron. 2019. Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Aff.* 98 (2019), 147.

[8] Umur A Ciftci, Gokturk Yuksek, and Ilke Demir. 2023. My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1369–1379.

[9] Michael A Ciranni and Arthur P Shimamura. 1999. Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25, 6 (1999), 1403.

[10] Dianne Cyr, Milena Head, Hector Larios, and Bing Pan. 2009. Exploring Human Images in Website Design: A Multi-Method Approach. *MIS Q.* 33, 3 (Sept. 2009), 539–566.

[11] Jelle Demanet, Kristof Dhont, Lies Notebaert, Sven Pattyn, and André Vandierendonck. 2007. Pixelating familiar people in the media: Should masking be taken at face value? *Psychologica belgica* 47, 4 (2007), 261–276.

[12] Amandeep Dhir, Puneet Kaur, Kirsti Lonka, and Marko Nieminen. 2016. Why do adolescents untag photos on Facebook? *Computers in Human Behavior* 55 (2016), 1106–1115.

[13] Matthew Dunne-Miles. 2021. Deepfakes, dead relatives and digital resurrection. https://theface.com/society/deepfakes-dead-relatives-deep-nostalgia-ai-digital-resurrection-kim-kardashian-rob-kardashian-grief-privacy Retrieved November 3, 2021.

[14] Passant Elagroudy, Mohamed Khamis, Florian Mathis, Diana Irmscher, Andreas Bulling, and Albrecht Schmidt. 2019. Can Privacy-Aware Lifelogs Alter Our Memories?. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313052

[15] Passant Elagroudy, Mohamed Khamis, Florian Mathis, Diana Irmscher, Ekta Sood, Andreas Bulling, and Albrecht Schmidt. 2023. Impact of Privacy Protection Methods of Lifelogs on Remembered Memories. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 35, 25 pages. https://doi.org/10.1145/3491102.3517520

[16] Rakibul Hasan, David Crandall, Mario Fritz, and Apu Kapadia. 2020. Automatically detecting bystanders in photos to reduce privacy risks. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 318–335.

[17] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173621

[18] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2019. Can privacy be satisfying? On improving viewer satisfaction for privacy-enhanced photos using aesthetic transforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[19] Eman T. Hassan, Rakibul Hasan, Patrick Shaffer, David Crandall, and Apu Kapadia. 2017. Cartooning for Enhanced Privacy in Lifelogging and Streaming Videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1333–1342.

[20] Yubo Hou, Xinyu Pan, Xinyue Cao, and Qi Wang. 2022. Remembering online and offline: The effects of retrieval contexts, cues, and intervals on autobiographical memory. *Memory* 30, 4 (2022), 441–449.

[21] Hakun Hukkelas. 2021. DeepPrivacy. https://github.com/hukkelas/DeepPrivacy. Retrieved November 3, 2021.

[22] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. 2019. DeepPrivacy: A Generative Adversarial Network for Face Anonymization. In *Advances in Visual Computing*. Springer International Publishing, 565–578.

[23] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/Off: Preventing Privacy Leakage From Photos in Social Networks. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security* (Denver, Colorado, USA) *(CCS '15)*. ACM, New York, NY, USA, 781–792. https://doi.org/10.1145/2810103.2813603

[24] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 781–792.

[25] JPEG. 2021. Content Authenticity Initiative. https://contentauthenticity.org/. Retrieved November 3, 2021.

[26] JPEG. 2021. JPEG Fake Media Standard. https://jpeg.org/items/20200803_fake_media.html. Retrieved November 3, 2021.

[27] Mohamed Khamis, Habiba Farzand, Marija Mumm, and Karola Marky. 2022. DeepFakes for Privacy: Investigating the Effectiveness of State-of-the-Art Privacy-Enhancing Face Obfuscation Methods. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. 1–5.

[28] kissclipart. 2021. Smiley Face Background. https://www.kissclipart.com/emojis-png-clipart-emoji-sticker-r1m975/. Retrieved November 3, 2021.

[29] Pavel Korshunov, Andrea Melle, Jean-Luc Dugelay, and Touradj Ebrahimi. 2013. Framework for objective evaluation of privacy filters. In *Applications of Digital Image Processing XXXVI*, Vol. 8856. International Society for Optics and Photonics, 88560T.

[30] Udo Kuckartz. 2016. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung (Grundlagentexte Methoden, 3., überarbeitete Auflage). *Weinheim: Beltz Juventa. Zugriff am* 9 (2016), 2017.

[31] Jari Kätsyri, Klaus Förger, Meeri Mäkäräinen, and Tapio Takala. 2015. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology* 6 (2015). https://doi.org/10.3389/fpsyg.2015.00390

[32] Karen Lander, Vicki Bruce, and Harry Hill. 2001. Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 15, 1 (2001), 101–116.

[33] Fenghua Li, Zhe Sun, Ang Li, Ben Niu, Hui Li, and Guohong Cao. 2019. Hideme: Privacy-preserving photo sharing on social networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 154–162.

[34] Yifang Li and Kelly Caine. 2022. Obfuscation remedies harms arising from content flagging of photos. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–25.

[35] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–24.

[36] ImageFitler Module. 2021. ImageFilter Module. https://pillow.readthedocs.io/en/3.0.x/reference/ImageFilter.html. Retrieved November 3, 2021.

[37] M. Mori, K. F. MacDorman, and N. Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine* 19, 2 (2012), 98–100.

[38] Kyle Murray and Gerald Häubl. 2011. Freedom of Choice, Ease of Use, and the Formation of Interface Preferences. *MIS Quarterly* 35 (12 2011), 955–976. https://doi.org/10.2139/ssrn.1698204

[39] MyHeritage. 2021. Deep Nostalgia. https://www.myheritage.com/deep-nostalgia Retrieved November 3, 2021.

[40] José Ramón Padilla-López, Alexandros Andre Chaaraoui, Feng Gu, and Francisco Flórez-Revuelta. 2015. Visual privacy by context: proposal and evaluation of a level-based visualisation scheme. *Sensors* 15, 6 (2015), 12959–12982.

[41] Chi-Hyoung Rhee and C Lee. 2013. Cartoon-like avatar generation using facial component matching. *Int. J. of Multimedia and Ubiquitous Engineering* 8, 4 (2013), 69–78.

[42] Daniel L Schacter. 1999. The seven sins of memory: insights from psychology and cognitive neuroscience. *American psychologist* 54, 3 (1999), 182.

[43] Peter Seddon and Min-Yen Kiew. 1. A Partial Test and Development of Delone and Mclean's Model of IS Success. *Australasian Journal of Information Systems* 4, 1 (1 1). https://doi.org/10.3127/ajis.v4i1.379

[44] Matthew Smith, Christian Szongott, Benjamin Henne, and Gabriele Von Voigt. 2012. Big data privacy issues in public social media. In *2012 6th IEEE international conference on digital ecosystems and technologies (DEST)*. IEEE, 1–6.

[45] Charles B Stone and Qi Wang. 2019. From conversations to digital communication: The mnemonic consequences of consuming and producing information via social media. *Topics in cognitive science* 11, 4 (2019), 774–793.

[46] Jose M. Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo Privacy Conflicts in Social Media: A Large-scale Empirical Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3821–3832. https://doi.org/10.1145/3025453.3025668

[47] Rashid Tahir, Brishna Batool, Hira Jamshed, Mahnoor Jameel, Mubashir Anwar, Faizan Ahmed, Muhammad Adeel Zaffar, and Muhammad Fareed Zaffar. 2021. Seeing is Believing: Exploring Perceptual Differences in DeepFake Videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[48] Shahroz Tariq, Sowon Jeon, and Simon S Woo. 2021. Am I a Real or Fake Celebrity? Measuring Commercial Face Recognition Web APIs under Deepfake Impersonation Attack. *arXiv preprint arXiv:2103.00847* (2021).

[49] Kurt Thomas, Chris Grier, and David M Nicol. 2010. unfriendly: Multi-party privacy risks in social networks. In *Privacy Enhancing Technologies: 10th International Symposium, PETS 2010, Berlin, Germany, July 21-23, 2010. Proceedings 10*. Springer, 236–252.

[50] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia. 2020. Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion* 64 (2020), 131–148.

[51] Kaveh Waddell. 2018. The impending war over deepfakes. https://www.axios.com/the-impending-war-over-deepfakes-b3427757-2ed7-4fbc-9edb-45e461eb87ba.html Retrieved November 3, 2021.

**Perceived Aesthetics**

| Obfuscating others | Original | Blurring | Pixelating | Masking | DeepFake | Avatar |
|---|---|---|---|---|---|---|
| Original | - | p < 0.001 | p = 0.001 | p < 0.001 | p = 0.002 | p = 0.001 |
| Blurring | p < 0.001 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Pixelating | p = 0.001 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Masking | p < 0.001 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 |
| DeepFake | p = 0.002 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 |
| Avatar | p = 0.001 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - |

**Perceived Privacy Protection**

| Obfuscating others | Original | Blurring | Pixelating | Masking | DeepFake | Avatar |
|---|---|---|---|---|---|---|
| Original | - | p = 0.001 | p < 0.001 | p < 0.001 | p < 0.001 | p < 0.001 |
| Blurring | p = 0.001 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Pixelating | p < 0.001 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Masking | p < 0.001 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 |
| DeepFake | p < 0.001 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 |
| Avatar | p < 0.001 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - |

**Perceived Integration of obfuscation within the photo**

| Obfuscating others | Original | Blurring | Pixelating | Masking | DeepFake | Avatar |
|---|---|---|---|---|---|---|
| Original | - | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p = 0.002 |
| Blurring | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Pixelating | p > 0.0033 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Masking | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 |
| DeepFake | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p = 0.002 |
| Avatar | p = 0.002 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p = 0.002 | - |

**Photo information sufficiency**

| Obfuscating others | Original | Blurring | Pixelating | Masking | DeepFake | Avatar |
|---|---|---|---|---|---|---|
| Original | - | p > 0.0033 | p = 0.002 | p = 0.001 | p = 0.003 | p = 0.002 |
| Blurring | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Pixelating | p = 0.002 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Masking | p = 0.001 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 |
| DeepFake | p = 0.003 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 |
| Avatar | p = 0.002 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - |

**Perceived Aesthetics**

| Obfuscating self | Original | Blurring | Pixelating | Masking | DeepFake | Avatar |
|---|---|---|---|---|---|---|
| Original | - | p < 0.001 | p = 0.001 | p < 0.001 | p = 0.002 | p < 0.001 |
| Blurring | p < 0.001 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Pixelating | p < 0.001 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Masking | p < 0.001 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 |
| DeepFake | p = 0.002 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 |
| Avatar | p < 0.001 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - |

**Perceived Privacy Protection**

| Obfuscating self | Original | Blurring | Pixelating | Masking | DeepFake | Avatar |
|---|---|---|---|---|---|---|
| Original | - | p = 0.001 | p < 0.001 | p < 0.001 | p < 0.001 | p < 0.001 |
| Blurring | p = 0.001 | - | p > 0.0033 | p = 0.001 | p = 0.002 | p > 0.0033 |
| Pixelating | p < 0.001 | p > 0.0033 | - | p = 0.0032 | p = 0.003 | p > 0.0033 |
| Masking | p < 0.001 | p = 0.001 | p = 0.0032 | - | p > 0.0033 | p > 0.0033 |
| DeepFake | p < 0.001 | p = 0.002 | p = 0.003 | p > 0.0033 | - | p > 0.0033 |
| Avatar | p < 0.001 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - |

# A Detailed Statistical Analysis

The details of the statistical analysis are in the next two pages.

**Perceived Integration of obfuscation within the photo**

| Obfuscating self | Original | Blurring | Pixelating | Masking | DeepFake | Avatar |
|---|---|---|---|---|---|---|
| Original | - | p = 0.002 | p = 0.001 | p = 0.001 | p > 0.0033 | p = 0.002 |
| Blurring | p = 0.002 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Pixelating | p = 0.001 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Masking | p = 0.001 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 |
| DeepFake | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 |
| Avatar | p = 0.002 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - |

**Photo information sufficiency**

| Obfuscating self | Original | Blurring | Pixelating | Masking | DeepFake | Avatar |
|---|---|---|---|---|---|---|
| Original | - | p = 0.003 | p = 0.002 | p > 0.0033 | p = 0.002 | p = 0.001 |
| Blurring | p = 0.003 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Pixelating | p = 0.002 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 | p > 0.0033 |
| Masking | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 | p > 0.0033 |
| DeepFake | p = 0.002 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - | p > 0.0033 |
| Avatar | p = 0.001 | p > 0.0033 | p > 0.0033 | p > 0.0033 | p > 0.0033 | - |


[52] Qi Wang. 2013. *The autobiographical self in time and culture.* Oxford University Press, Oxford, UK.

[53] Qi Wang. 2021. The cultural foundation of human memory. *Annual review of Psychology* 72 (2021), 151–179.

[54] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. "I regretted the minute I pressed share" a qualitative study of regrets on Facebook. In *Proceedings of the seventh symposium on usable privacy and security.* 1–16.

[55] Mika Westerlund. 2019. The emergence of deepfake technology: A review. *Technology Innovation Management Review* 9, 11 (2019).

[56] Mika Westerlund. 2019. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review* 9 (11/2019 2019), 40–53. https://doi.org/10.22215/timreview/1282

[57] Leslie Wöhler, Satoshi Ikehata, and Kiyoharu Aizawa. 2024. Investigating the Perception of Video Anonymization Techniques in 360 {\deg} Videos. *arXiv preprint arXiv:2408.04844* (2024).

[58] Aimin Xu, Shitao Fang, Huan Yang, Simo Hosio, and Koji Yatani. 2024. Examining Human Perception of Generative Content Replacement in Image Privacy Protection. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 777, 16 pages. https://doi.org/10.1145/3613904.3642103

[59] Xiping Xiong, Akira Chinomi, Takashi Kishimoto, Noboko Setsu, Yoshiaki Ishi-Ito, and Noboru Babaguchi. 2008. Privacy protecting visual processing for secure video surveillance. In *2008 15th IEEE International Conference on Image Processing*. IEEE, 1672–1675.