A deepfake is the use of artificial intelligence (AI) to create hyper realistic photos or videos by superimposing a person's face onto another's body (Chawla, 2019). This emerging phenomenon has been spreading over the past few years as advancements in technology have made deepfake creation more accessible to the general public and as a result more deepfakes have been released on social media platforms. In some cases, these deepfakes have resulted in irreparable harm to their targets who are left with little legal recourse. This proliferation of deepfakes is an example of the all too common problem of technology advancing faster than the laws meant to govern its usage.

Due to the novelty of deepfake technology, there is a lack of research on deepfakes in the legal system. This study looks at the community's views on the legality of deepfakes. Deepfakes can be made for a variety of reasons. The focus of this research is on individual's making deepfakes, without permission, of individuals at different levels of public exposure (celebrity, politician, social media influencer, private citizen) and the type of deepfake (illicit material, satire) made. The level of public exposure variable examines view on legality based on the public's access to the person. While the type of deepfake examines where the community draws the line on what type of deepfake should or should not be allowed to be made. Two facets of legality were used: whether the deepfake is allowed to be created and penalty given to creators of the deepfake.

It was hypothesized that, in general, participants will be less accepting of illicit deepfakes than satirical deepfakes. It was also hypothesized that given their choice to be more public, deepfakes made of celebrities, politicians, and influencers will be deemed more acceptable than private citizens.

Methods

Participants

Participants were recruited via the university's research management system. Participants consisted of 137 college students at a moderate sized university in the southern part of the United States in undergraduate psychology courses. Participants were predominately female (81.8%, 17.5% male, and .7% non-binary) with an average age of 19.10 (SD = 1.82) years old. The racial/ethnic identity of the participants was predominately Caucasian (78.8%), with Black (9.5%), Hispanic (3.6%), Asian (2.9%), multi-Racial (2.9%), and other (.7%).

Materials and Procedures

After completing the consent form, an introduction on the study was given explaining the definition of a deepfake, showing an example of a deepfake, and supplying instructions for the survey. To emphasize how realistic deepfakes look, the example consisted of side-by-side before and after photos along with the original source material for the deepfake target (i.e. person who's face was morphed onto the body). To avoid any contamination, a non-celebrity neutral deepfake was utilized.

Following the instructions, participants were presented with eight different scenarios within subjects. Written scenarios were based on the type of deepfake (illicit, satire) and the target's

level of public exposure (celebrity, politician, social media influencer, private person). An example of an illicit material scenario is, “A person created a deepfake placing a celebrity's face, taken from their social media account, onto a body of a person engaging in illicit behavior on an adult website”. Meanwhile, an example of a satire scenario is, “A person created a deepfake placing a celebrity’s face, taken from their social media account, onto their own to mock the celebrity’s characteristics and behaviors”. For each level of public exposure, the word “celebrity” was simply replaced with the other three levels.

After each scenario, a questionnaire was given to test legality including both whether that type of deepfake should be allowed and the penalties a participant believes the creator should receive. The Allowed condition was shown as a 7-point Likert scale from very strongly disagree (0) to very strongly agree (6). A scale that increased with severity was given for the Penalties condition (none, account suspended, verbal reprimanded, probation, jail time).

Results

Separate repeated measures ANOVA was performed to examine the effects of the type of deepfake on whether individuals should be allowed to create that type of deepfake (Allowed, $M = 0.93$, $SE = 0.08$) and how harsh a penalty creators should receive (Penalty, $M = 1.79$, $SE = 0.07$). The analysis for Allowed was significant, $F(7, 910) = 54.27$, $p < .001$, partial $\eta 2 = 0.29$. Pairwise comparisons revealed in general participants found creating the deepfake for illicit purposes was significantly less acceptable than for satire purposes (all $ps < .001$). Also of interest, using a private person for satire purposes was also deemed significantly less acceptable than a celebrity ($p = .003$) or politician ($p = .008$) being utilized for satire purposes. The analysis for Penalty was also significant, $F(7, 910) = 54.27$, $p < .001$, partial $\eta 2 = 0.29$. Pairwise comparisons were nearly identical to those for Allowed with creating a deepfake for illicit purposes resulting in significantly higher penalties than satire purposes, (all $ps < .001$). Again, using a private person for satire purposes resulted in significantly higher penalty than for celebrity ($p = .028$), a politician ($p = .009$) and social media influencer ($p = .036$).

Discussion

The results overall confirmed the hypothesis. Participants viewed the deepfakes in the Illicit condition as less acceptable than the ones for the Satire condition and recommended a higher penalty. Furthermore, creating a deepfake using a private citizen was deemed less acceptable and worthy of a harsher punishment than the more public figures. One possibility for this is participants empathized more with the private person than the public figures due to having the same public status. Another possibility is participants viewed public figures as voluntarily making their images accessible compared to the private individual. Of note, overall means for deepfake acceptance was extremely low across all conditions, indicating a rather strong dislike of deepfakes being made. These findings could be used to guide public policy regarding deepfakes as well as future research into their legality.

Abstract

Deepfakes, using artificial intelligence to create hyper realistic photos or videos by superimposing a person's face onto another's body, is an emerging issue with little legal precedent. The current research examines participant opinions regarding acceptable creation of deepfakes. Participants rated acceptability of several scenarios in which the goal of the deepfake (satire or illicit) and target's level of public exposure (celebrity, politician, influencer, private citizen) were varied. Results indicated illicit deepfakes were significantly less acceptable than satire deepfakes, and private citizens were less acceptable targets of deepfake creation. These findings could help guide future public policy concerning deepfakes.