Does Privacy on Social Media Exist?

The Newest Threat of Deepfakes to Internet Privacy

Nina Squicciarini

HONR 420: Honors Thesis

Dr. Wenjing Xie

Table of Contents

Abstract

Social media has become a regular part of many people's everyday lives. Most social media users are unaware of the dangers and threats that come with putting your personal information on the Internet. The newest threat to Internet privacy is the artificial intelligence (AI) program, Deepfakes. This program uses a collection of images to create fake videos of people. Once Deepfake technology is more accessible, there will be plenty of opportunities for anyone to use people's images from social media to create fake videos. There is no privacy protection on social media that would help prevent this situation, and there are no laws to support victims from this new spread of misinformation. In this study, an online experiment was conducted with four different conditions of varying proximity and efficacy involving the threat of Deepfakes. A total of 203 Marist students participated and the majority had never heard of this AI program. The respondents already did not trust Instagram and Facebook to protect their information, but they will continue to actively use these platforms. The participants plan to change their future behavior by restricting certain information they post on social media and limiting the people who follow their accounts. Many felt that Deepfakes are a distant threat and were unsure of how it could impact them. Currently, people value the instant gratification of social media over their own privacy on the Internet.

Introduction

Artificial intelligence (AI) is continuing to develop each and every day. The newest AI program to fear is called Deepfakes. This program scans images of people to create a digital version of them. This digital version of the person can be manipulated and used to create any video. These Deepfake videos are becoming more difficult to identify as the program technology is constantly being enhanced. Currently, politicians and celebrities have been victims of this program and its power to spread false information. Soon the next victim will be normal everyday people. Some viral Deepfake videos include former President Barack Obama, Mark Zuckerberg, and Jennifer Lawrence (Kietzmann et al., 2020).

The name Deepfakes came from the machine learning method called deep learning. Deep learning involves artificial intelligence processes data and "learning" how to make decisions. A popular form of AI that uses deep learning is Google Translate. Google Translate was created through the upload of mass amounts of data from different languages. Google Translate is then able create predictive algorithms and make decisions using the inputted data. AI technology has been developed and advanced to now make videos, which is where Deepfakes come into play. The Deepfakes programming involves uploading images into the program. With enough images and varying facial expressions, the program is able to create a digital version of a person. For example, if a person uploaded a large amount of pictures of Katy Perry. That person could now film a video as Katy Perry and make "her" say or do whatever they'd like. Celebrities and politicians have been easy targets due to the high quality of pictures of them available on the Internet. Another place on the Internet with an endless amount of pictures is social media.

As artificial intelligence programs continue to develop and grow, they will become more accessible to people. Over the past few years, Deepfakes went from a program that required an

expert programmer, to a feature that is being incorporated in apps. Soon this program will be available for anyone to use, and for any purpose. Social media has become a regular part of people's lives. Instagram is one of the most popular platforms and users will regularly post pictures of themselves to share with their friends and family. This makes social media accounts the perfect place to steal pictures to upload into a Deepfake program. Everyone with a social media account is a potential victim of this new AI program. There is very little research on this new AI program due to how new it is, but there is plenty of research about the lack of privacy on social media. There are currently over 3.6 billion social media users and still no legal protection from companies such as Instagram and Facebook (Fallis, 2020). These social media companies are actively selling information to unauthorized parties everyday. This will become a larger issue when people become more aware of where their information is going and how it can be used to spread misinformation. Deepfakes is an issue that people should be aware of in order to gain control of their online information and privacy. This study focuses on how college students perceive the threat of Deepfakes and their Internet privacy when there is a variation in proximity and efficacy.

Literature Review

Deepfakes

In December 2017, the first Deepfake appeared on the social news platform, Reddit. Deepfakes make people on the Internet, and their information, very vulnerable. Researchers have discovered that there are two extremes with Deepakes. Some people view deepfakes as a privacy threat on the Internet, while others view it as a new technological development that will enhance films (Yadlin-Segal & Oppenheim, 2020). Deepfakes is very new technology and this makes it

difficult to clearly see which direction this development will follow. It is important to recognize the potential threats that Deepfakes have created. This will become an issue to the many people on the Internet who actively use social media. With Deepfakes on the rise, it will be essential to regulate and moderate content on social media (Yadlin-Segal & Oppenheim, 2020). Deepfakes will continue to develop and become more difficult to distinguish. This creates a privacy issue for everyone with pictures, videos, and information on the Internet and social media

The Deepfakes program has the technology to manipulate audio, videos, and pictures. The plan to manage the risk of Deepfakes involves recording the original video in order to deny fake videos, exposing Deepfakes in the early stages, gain legal protection against Deepfakes, and making people aware of the risk (Kietzmann et al., 2020). Currently, Deepfakes have been used for comedy videos, as well as, spreading misinformation about politicians (Yadlin-Segal & Oppenheim, 2020). This technology will soon become an issue for everyday people. Deepfakes impacted Scarlett Johannson, Katy Perry, Barack Obama, Donald Trump, and more (Kietzmann et al., 2020). It is essential to recognize the harm this artificial intelligence program can cause, and how people can be proactive to protect their private information and images.

Deepfakes use machine learning to create the most realistic digital version of a person. Machine learning has made it easy for people to create fake videos of others (Fallis, 2020). As Deepfakes continue to develop, the program requires less data to create these fake videos (Kietzmann et al., 2020). In 2019, China created a free, popular app called Zao. This app allows users to put their face on characters in movies and television shows. The only requirements to create these fake videos are a few selfies with different facial expressions. Some users sought out a program called Deepmind's WaveNet, to create realistic speech from inputted text (Kietzmann et al., 2020). These videos are easy to create and could easily be used for malicious purposes in

the future. Deepfakes are not perfect, but their technology is improving and it will become more difficult to distinguish between real and disingenuous videos (Fallis, 2020). This artificial intelligence program is becoming more widely available and this means that everyday people will have access to create these videos.

Currently, there are no legal consequences or protection from Deepfakes. In the future, forensic scientists may be about to identify programs on laptops and mobile phones, but this is not guaranteed (Venema & Geradts, 2020). Judges and juries would have to be educated on the topic and need an overwhelming amount of evidence to ensure that the video is fake. People are more likely to believe videos that come from a trustworthy source, such as legitimate news outlets (Fallis, 2020). As Deepfake videos become difficult to identify with the plain eye, it will be extremely difficult to clearly determine whether a video is real or digitally altered. It is essential that programs are utilized to look for ways to identify fake videos in the software to provide concrete evidence that a video is not authentic.

Social cognitive theory

The social cognitive theory has been applied primarily to research involving health information. Many people post personal information on social media, and this includes health information. People want to be a part of this online community, but it also presents a threat when their personal information is shared. Social media provides a connection without meeting face-to-face (Lin & Chang, 2018). The sole purpose of social media is to exchange information, and this could be through images or text. The benefits outweigh the privacy risk for social media users.

Social cognitive theory specifically focuses on behavior change. This theory has been applied to health issues with high self-efficacy (Riley et al., 2016). For example, smoking and weight gain can be controlled through changes in behavior. Social media creates a community where people can share information which can help others make difficult behavior changes (Lin & Chang, 2018). Self-efficacy is associated with the expected outcome from these decisions. The perceived social support and self-efficacy influence the relationship people have with certain barriers (Riley et al., 2016). If people have social support and the ability to control a certain obstacle, they are more likely to make that behavior change to achieve their desired outcome. It is important that people have the necessary information and support to create positive change in their life.

A study was conducted with participants that have MS and the experiment involved them self-reporting their assessments. This procedure allowed researchers to study self-efficacy. Participants with Multiple Sclerosis (MS) were encouraged to exercise and self-report their progress. The self-report resulted in an increase in exercise goal setting, planning and benefits (Uszynski et al., 2018). These inactive adults were more likely to exercise when they self-reported. This demonstrates the influence that self-efficacy had on the participants. Self-efficacy applies to exercise because this is a factor that you can control. Social relationships also had an impact on individuals' self-efficacy (Uszynski et al., 2018). In this study, participants were encouraged by the people around them, as well as, the researchers checking in on them through the phone. An individual's community and direct network can influence your self-efficacy (Riley et al., 2016).

Human behavior is constantly changing. As technology continues to develop, human behavior will continue to change as well (Ratten V. & Ratten H., 2007). Human behavior is

influenced by the attitudes of others and society deem acceptable. Technology is becoming more advanced, and the more that people post information on the Internet, the less control they will have over their information. Social media has become a place where people feel comfortable sharing personal information with others. The relationships that are established determine how much information people are willing to share. Social media has created a community where people feel comfortable sharing private information such as their health (Lin & Chang, 2018). This exchange involves both sharing and seeking information. The online community that social media creates, influences people's decisions and relationships (Uszynski et al., 2018).

Media is an essential part of communication. Media can spread awareness and also influence people's behavior. Television, the news, and the Internet all impact individuals' behavior and actions (Ratten V. & Ratten H., 2007). Technology is making it easier for people to communicate online and share personal information. Sharing personal information is a risk that people are willing to take to gain their desired social connection (Lin & Chang, 2018).

Construal level theory

The construal level theory focuses on the psychological distance of a certain threat. Brügger, Morton, and Dessai conducted two experiments to analyze how individuals' motivation is influenced by their perceived distance of climate change. Participants that were further from the image of climate change were more skeptical while making decisions, while those that were close to the effects of climate change made their decisions based on fear (Brügger et al., 2016). Also, participants were more concerned with events that would specifically impact and influence them. This shows how people are more interested and aware when they will be directly impacted. This relates to the privacy risks and threats on the Internet. The Internet and social

media are constantly collecting data using third party sources, and users are not concerned with this issue because it's viewed as a distance problem (Liberman et al., 2007). This study will analyze how people view their Internet privacy depending on how close the problem is.

There is no clear answer as to whether it is better to make a decision with a distanced perspective or a proximal perspective. Each situation is different and involves various factors that influence an individual's decisions. Individuals that are psychologically closer to the issues and more likely to take action compared to those who are psychologically further (Liberman et al., 2007). A main factor in these decisions is the risk involved. People are more likely to make behavioral changes when they fear the consequences of a certain issue (Brügger et al., 2016). People tend to be excessively sensitive to risk when their money is involved; this includes buying insurance and investing in stocks (Liberman et al., 2007).

Evidence shows that there is an indirect relationship between psychological distance and mental representation (Henderson et al., 2011). Those with higher mental representation are capable of perceiving more distance. This demonstrates how when an individual has a positive view of someone, they are able to see their relationship is a further distance. The mental representation of this said person will influence the individual's behavior and decisions regarding the future. Psychological distance is how far an individual perceives an event or person (Liberman et al., 2007). The way people view social media is impacting their decisions about their privacy. People that actively use social media have a positive view of the platform and this encourages them to continue using these applications (Henderson et al., 2011). Getting likes on social media releases dopamine in the brain and people have become addicted to this positive attention. Many people will plan in advance when they will post on social media and specifically

go to events to post on these platforms. With the social support of their peers, it becomes expected to actively post on social media.

Attracting a consumer involves their psychological distance. The main goal for marketing and media is to convince consumers to use their product or service (Fiedler, 2007). The goal is to influence consumers for the long-term goal so that as distance decreases, their appeal to your brand increases. This creates loyal consumers that continue to support your business (Fiedler, 2007). This relates to social media because it is constantly changing to appeal to their audience. Many people now work as influencers on social media as a career, and this has attracted more people and businesses to the platform. An individual's level of mental representation of a certain person or organization influences their decisions (Henderson et al., 2011). The entertainment and value of social media blinds people from truly recognizing potential threats. This research led to the following research questions:

RQ1: Does proximity and efficacy influence how people view the threat of Deepfakes?

Trust on social media

Trust is very complex and is established through several layers of communication. Trust is based on the idea of making decisions that help maintain long-term relationships involving collaboration (Cho et al., 2015). People are more likely to self-disclose with people they trust and already have an established relationship with (Vogel & Wester, 2003). There are four different categories of trust: communication trust, information trust, social trust, and cognitive trust (Cho et al., 2015). Communication trust involves response time and the quality of service. Information trust is based on the credibility of the network and social trust focuses on the source's reliability. Lastly, cognitive trust involves the capability of processing the information

(Cho et al., 2015). Different fields value and measure trust differently. Those in social media measure the trust that the users have in the platform and their trust in the quality of service being provided. Also, people are more likely to trust outlets and platforms that are deemed socially acceptable by the majority (Turcotte et al., 2015). Trust involves taking a risk and being uncertain of the outcome. Trust involves several different factors and layers of communication.

Media is becoming more complex and this is making it more difficult for outlets to retain an audience (Turcotte et al., 2015). Many people have turned to getting their news from social media. People have established a form of trust in social media, and this has had an impact on the success of news outlets. Logical trust and emotional trust influences individual's decisions regarding where they get their information (Cho et al., 2015). The research concluded that people do not trust national news outlets and this has led people to seek more information from social media. Social media consists of both opinion leaders and opinion seekers (Turcotte et al., 2015). These roles flow interchangeably depending on the issue being discussed. Individuals are also more likely to share personal views and information to their trusted communities (Bazarova & Choi, 2014). People trust those in their community with similar views. If a trusted opinion leader supports a specific outlet, people are more likely to trust this outlet as a source (Turcotte et al., 2015). Social media allows people to seek information directly from the trusted opinion leader's accounts, and this prevents the news from interpreted information inaccurately. People trust social media more because it provides more raw information from individuals.

Self-disclosure involves revealing private and personal information to others (Bazarova & Choi, 2014). This type of communication is intentional and requires a certain degree of trust. Disclosing information allows people to strengthen their connection with others and fulfills the social need of communication (Bazarova & Choi, 2014). Studies showed that people had a better

attitude once they self-disclose personal information to others (Vogel & Wester, 2003). Even though self-disclosure can be beneficial, it also makes people vulnerable and presents risks. This is due to the lack of control once this information is shared with others. A way to gain some control is to specifically share this information with trusted individuals (Bazarova & Choi, 2014). On social media, people have the freedom to share information everyday. Some people have public accounts while others have private accounts in an attempt to control who sees the information they share.

Individual's trust is divided into two sections: logical trust and emotional trust (Cho et al., 2015). Logical trust is based on experience, credibility, and rationality, while emotional trust is based on fear, expectation, and regret (Cho et al., 2015). Social media has provided an outlet for people to show more of who they are. This has led to personal connections, but it has also led to personal information being used in malicious ways. It is important for social media users to be aware of the consequences and risks of self-disclosure on the Internet (Bazarova & Choi, 2014). Disclosing information is a "risk-taking behavior" (Vogel & Wester, 2003). For many people, disclosing personal information is worth the risk when there is the opportunity to expand relationships with others on social media.

Privacy protection on social media

Even though teenagers have grown up on the Internet connecting with people, they still want privacy. Of American teenagers, 95% use the Internet and 85% are on social media (Marwick & Boyd, 2014). This study acknowledged that teenagers prefer to share only general information about themselves such as their names, hobbies and places they like. Teenagers keep personal information private and many are aware of the consequences of social media. When you

put information about yourself online, it creates an opportunity for people to "cyber-bully" and hurt others from behind their computer screen. In the U.S., 72% of teenagers have experienced online harassment (Lwin et al., 2012). Online harassment has impacted teenagers' mental health, self esteem, and quality of life. People have been taking precaution online due to the extensive amount of harassment. People have been using defense technology and withholding information to protect their identity online and remain safe on the Internet (Lwin et al., 2012). This behavior is dangerous and the development of Deepfakes makes this threat even more serious.

Each social media platform has different settings that are available to provide privacy for the users (Marwick & Boyd, 2014). Even if individuals have control over their own social media accounts, they cannot control what their friends or others are posting about them. The concept of privacy is constantly changing as technology and the Internet continue to develop (Hadar et al., 2017). Teenagers are often concerned about the way their social media profiles will look to college recruiters and this motivates them to have a presentable image online (Marwick & Boyd, 2014). This type of control is limited and there is no protection regarding the other information about you online that is posted by others. The lack of control on the Internet makes it difficult for anyone to protect their information.

Software developers are meant to incorporate the privacy design of a new technology in the beginning stages of project development. When developing new technology, it is essential for developers to create and design privacy (Hadar et al., 2017). The design of privacy all depends on the interpretation of the developers. This gives developers the power to decide what information is being collected from the users, as well as, how to protect certain information.

Social media makes it difficult to keep information private even when all precautions are taken. Facebook has a facial recognition algorithm that suggests to "tag" people in posts that are

not theirs (Marwick & Boyd, 2014). This prevents users from protecting their images on the Internet. Participants in the study felt that privacy is not a technical concern, but rather a social concern (Hadar et al., 2017). Decisions involving privacy and based on current social norms. As technology advances, designers will focus on usability and functionality goals, and it is important that privacy is also considered essential. Technological developments such as, the Internet, provide many benefits but there is also the risk of online harassment.

Social media has become an important tool for sharing information. Many people actively use social media and this number is continuously growing. Privacy is a concern on social media, especially when third party programs obtain users' information for businesses (Kumar et al., 2016). Privacy on social media also depends on the specific platform and the privacy settings they offer. The extensive amount of people using social media, these platforms can experience technical difficulties (Hadar et al., 2017). Technical difficulties and glitches within the social media platform's programming can lead to user's information being scattered and the risk of a virus (Kumar et al., 2016). It is important that user's create strong passwords and change their passwords often to create the most secure account to protect their private information on social media. This led to the following research questions:

RQ2: Do the benefits of social media outweigh the privacy risks?

Methods

Design

This study consists of an online experiment with four different conditions to analyze the level of fear associated with Deepfakes and Internet privacy. Each condition had a different variation of proximity and efficacy using a Likert scale. All of the questions in the experiment

were the same, and the only difference was the condition created for each. The four conditions provided essential information on how people react to a threat based on the distance and control over the issue. The experiment was sent in the form of a questionnaire using Google Forms. This allowed the maximum reach and provided the maximum amount of responses. Online questionnaires are timely and participants can complete them on their mobile devices or computer.

Sampling

The study conducted includes Marist students ages 18-22 of varying demographics. The sample size includes 203 people, which breaks down to approximately 50 people per condition. This amount of responses provided the informative data and resulted in identifiable patterns. The Google form link was sent to students through email, text messages, as well as, Instagram and LinkedIn direct messages. Students in classes were offered extra credit as an incentive to complete the questionnaire.

This sample was chosen to provide the most controlled and effective study. Focusing specifically on college students was essential. This age group heavily relies on social media, and this is an integral part of the experiment. Students of different genders, household income, and ethnicity were included in this sample to gain an accurate representation of the population. Non-probability sampling was utilized to collect an appropriate amount of results for the study.

The participants were recruited through systematic sampling and convenience sampling. A list of 170 participants at Marist College was generated and randomized. This list was then divided into four groups to provide the most effective and unbiased results. In addition to this,

convenience sampling was used in communication classes to reach more students. In these classes, extra credit was offered as an incentive and this resulted in a total of 203 responses.

Measurement

Independent Variables:

Social media experience was measured through nine questions using Lin and Chang's study (2018) involving social media and the exchange of health information. The questions included how often each social media platform is used, as well as, the amount of friends and followers the respondent has. The social media platform answers were measured using a Likert Scale (1=Never, 2 = Several times a year, 3 = Several times a month, 4 = Several times a week, and 5 = Several times a day) and the number of followers were measured using a Multiple Choice Grid (0-100, 101-300, 301-500, 501-700, 701-900, and 900+).

Knowledge of Deepfakes was measured through one question. This question was adapted from Yadlin-Segal's (2020) study about the new developments of Deepfakes and the lack of regulation on social media. The question asked about the respondent's familiarity with Deepfakes. The answer was measured using a Likert Scale (1= I never heard of it and 5= I am very familiar with it).

Privacy concern for personal information was measured through five questions from the Hadar et al. (2017) study that analyzed online applications' privacy settings created by software developers. The questions included were respondents' comfort level with third parties sharing their information, concerns about their personal information, fear level about information being

safely stored, the misuse of their personal information, and whether they believe that companies are sharing their information without permission. These answers were measured on a Likert Scale (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree).

Dependent Variables:

Trust on social media was measured through three questions. These questions were adopted from Turcotte et al. (2015) research that involved the impacts of trusting information on social media and the impact this has on audiences. The questions included asked respondents if they view Facebook and Instagram as trustworthy, whether users trust social media companies to protect their personal information on their platforms, and if these social companies are successful in protecting this information. The answers we measured using a Likert Scale (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree).

Skepticism of Deepfakes was measured through five questions modified from Brügger et al. (2016) study regarding impact of the construal level theory. These questions included whether Deepfakes are seen as a threat by experts, their opinion on the media's response, the reliability of the evidence, if they believe this threat applies to them, and if they believe Deepfakes are a real risk. The respondents' answers were measured using a Likert Scale (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree).

Likelihood of self disclosure was measured through ten questions from Vogel and Wester's (2003) study about the risks of self-disclosure. The questions asked were whether they would post their home address, photos, videos, relationship status, real name, email address, cell phone

number, Instagram user ID, religion, and personal interests on social media. The answers were measured using a Likert Scale (1=Very Unlikely, 2=Unlikely, 3=Unsure, 4=Likely, and 5=Very Likely).

Privacy protection behavior was measured through eight questions altered from the Lwin et al. (2012) study involving adolescents and online harassment. The questions about social media were whether the participant would untag photos/videos of themselves, delete information, unfollow or unfriend people, make accounts private, make status updates private, remove apps, restrict information, and if they would install protection software. These answers were measured using a Likert Scale (1=Very Unlikely, 2=Unlikely, 3=Unsure, 4=Likely, and 5=Very Likely).

Perceived susceptibility was measured through eight questions from the Lwin et al. (2012) study. This included how likely respondents felt they would receive hate emails, online threats, sexual remarks online, someone pretending to be them, someone sharing their personal information with malicious intentions, someone posting personal photos/videos of them to intend harm, and someone spreading rumors about them online. The answers were measured on a Likert Scale (1=Very Unlikely, 2=Unlikely, 3=Unsure, 4=Likely, and 5=Very Likely).

General efficacy was measured through three questions from the Lwin et al. (2012) study. This included whether people believe it is possible to reduce privacy risks on the Internet if everyone does their part, individual behavior change, and online privacy protection methods. The answers were measured using a Likert Scale (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree).

Personal efficacy was measured through three questions adapted from the Lwin et al. (2012) study. This consisted of whether participants are able to protect their privacy on the Internet, if it is easy for them to reduce the online privacy risks, and if they are capable of adopting measures to protect their privacy online. These answers were measured through a Likert Scale (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree).

Behavioral intentions were measured through three questions modified from the Lwin et al. (2012) study. The questions about whether they would limit their personal information on social media, if they would reply to strangers on the Internet, and if they would disclose personal data on public websites. The answers were measured through a Likert Scale (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree).

Procedure

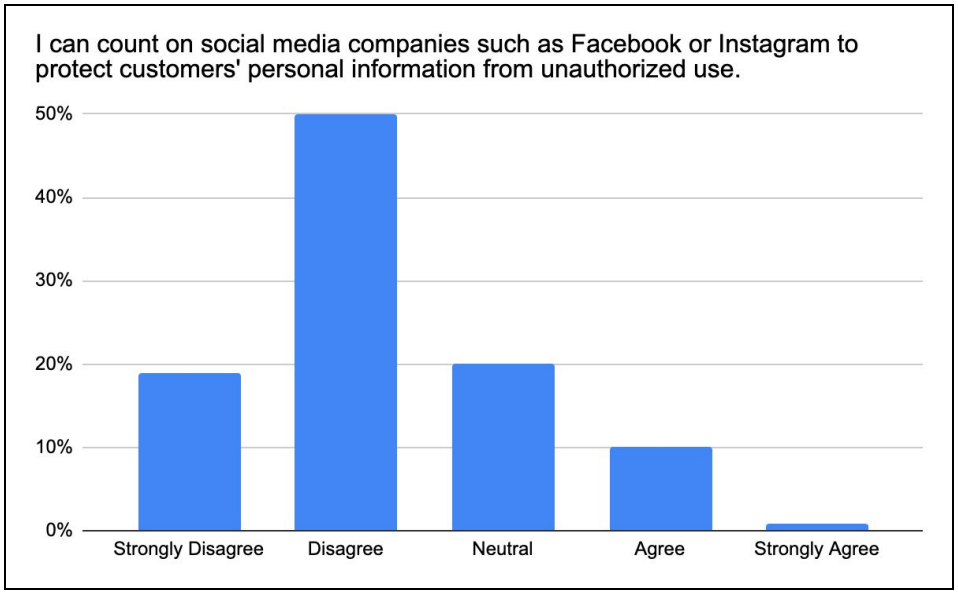
The questionnaires were sent to participants November 2nd-6th, 2020 using Google Forms. Google Forms is a free online survey application that was essential to reach participants during COVID-19. Each question required an answer and participants were informed that the questionnaire would take up to 10 minutes. Participants were also informed to complete the questionnaire as soon as possible in order to achieve the most accurate analysis of data. All participants were sent the link to the questionnaire through email, text message, as well as, Instagram and LinkedIn direct message. The online questionnaires were convenient and could be completed on any computer or mobile device. Approximately 250-300 people were sent the questionnaires in order to receive the maximum amount of responses. The questionnaire was

divided into two pages. The first page had general questions about the participant's social media usage and whether they have heard of Deepfakes. The second page went more into depth on Deepfakes and its potential threat, and ended with demographic questions.

Results

Participants were asked to respond to the following statement, "I can count on social media companies such as Facebook or Instagram to protect customers' personal information from unauthorized use." Of the 203 respondents, 19% strongly disagreed, 50% disagreed, 20% were neutral, 10% agreed, and 1% strongly agreed (Figure 1). The respondents were active on Instagram and Facebook, 91% use Instagram several times a day and 34% use Facebook several times a day. On Instagram, 61% of respondents had 900+ followers and 52% follow 900+ people. Respondents had a variety of answers regarding their Facebook friends, 23% had 301-500 friends and 22% had 101-300 friends. Social media is a part of college students' everyday lives and they are aware that their information is at risk on the Internet.

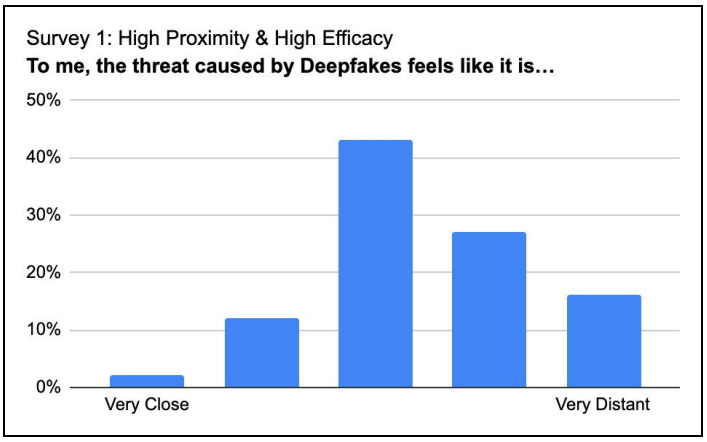
Figure 1:

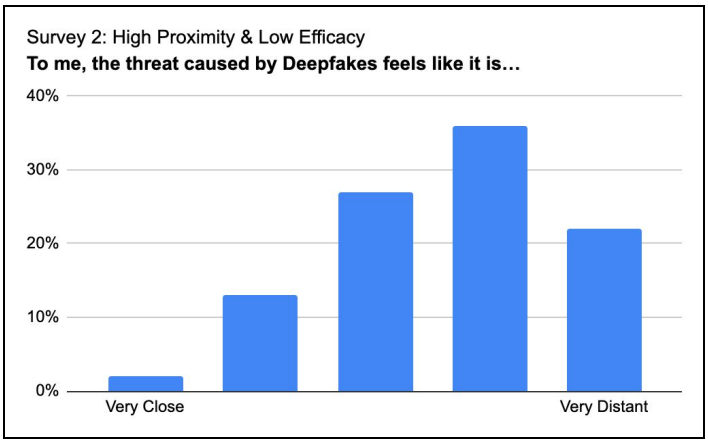


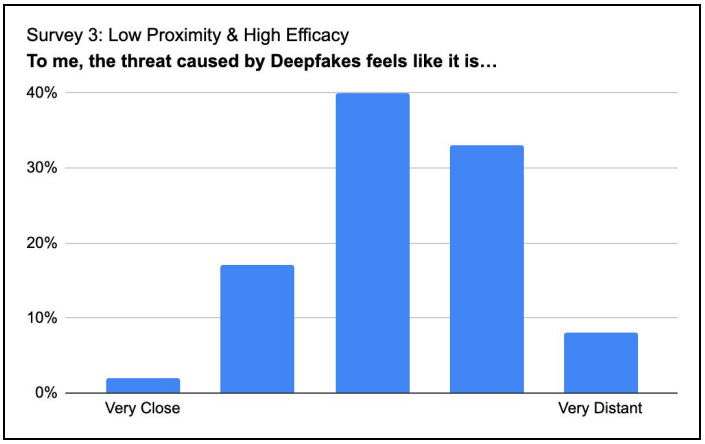
RQ1: Does proximity and efficacy influence how people view the threat of Deepfakes?

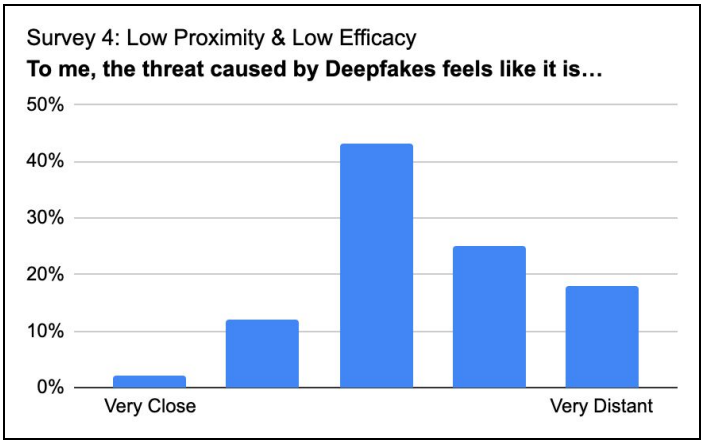
Many participants taking in the experiment had little knowledge about Deepfakes and 61% have never heard of the program before. Respondents were asked to rank how close the threat to Deepfakes feels to them (Figure 2). Each condition experienced different results due to the varying levels of proximity and efficacy. Overall, most of the respondents felt unsure about the situation (condition one 43%, condition two 27%, condition three 40%, & condition four 43%) or viewed Deepfakes as more of a distant threat (condition one 43%, condition two 58%, condition three 41%, & condition four 43%). A total of 43% of people agreed to the following statement, “I am uncertain if the risks caused by Deepfakes are happening to me.” People are unaware of the specific threats on the Internet that their personal information could be used for. Many participants throughout the experiment viewed Deepfakes risks as a real problem (condition one 47%, condition two 44%, condition three 35%, & condition four 47%).

Figure 2:


Survey 1: High Proximity & High Efficacy
To me, the threat caused by Deepfakes feels like it is…
50%
40%
30%
20%
10%
0%
Very Close
Very Distant


Survey 2: High Proximity & Low Efficacy
To me, the threat caused by Deepfakes feels like it is…
40%
30%
20%
10%
0%
Very Close
Very Distant


Survey 3: Low Proximity & High Efficacy
To me, the threat caused by Deepfakes feels like it is…
40%
30%
20%
10%
0%
Very Close
Very Distant


Survey 4: Low Proximity & Low Efficacy
To me, the threat caused by Deepfakes feels like it is…
50%
40%
30%
20%
10%
0%
Very Close
Very Distant

RQ2: Do the benefits of social media outweigh the privacy risks? Figure 3 demonstrates that across all conditions that respondents are now likely to restrict the types of information that they share with apps on social media (likely/very likely: condition one 75%, condition two 80%, condition three 80%, & condition four 69%). All respondents said they would unfriend or unfollow people on social media (likely/very likely: condition one 55%, condition two 58%, condition three 73%, & condition four 67%). Each condition had different results when asked if they would untag photos/videos of themselves on social media (condition one 43% disagree, condition two 33% disagree / 33% agree, condition three 33% agree, & condition four 33% unsure). It was evident that the majority of respondents would now delete information from their current social media (likely/very likely: condition one 55%, condition two 36%, condition three 56%, & condition four 51%).

In Figure 4, the results show how the majority of participants will continue to post photos/videos on social media. For photos the results were the following: condition one 92%, condition two 93%, condition three 96%, & condition four 86% responded likely/very likely. For videos the results were the following: condition one 88%, condition two 83%, condition three 83%, & condition four 70% responded likely/very likely. Participants were asked to respond to the following statement, “Taking actions to reduce online privacy risks is easy for me.” For condition one 47% agreed, condition two 53% were unsure, condition three 42% unsure, & condition four 53% agreed.

Figure 3:

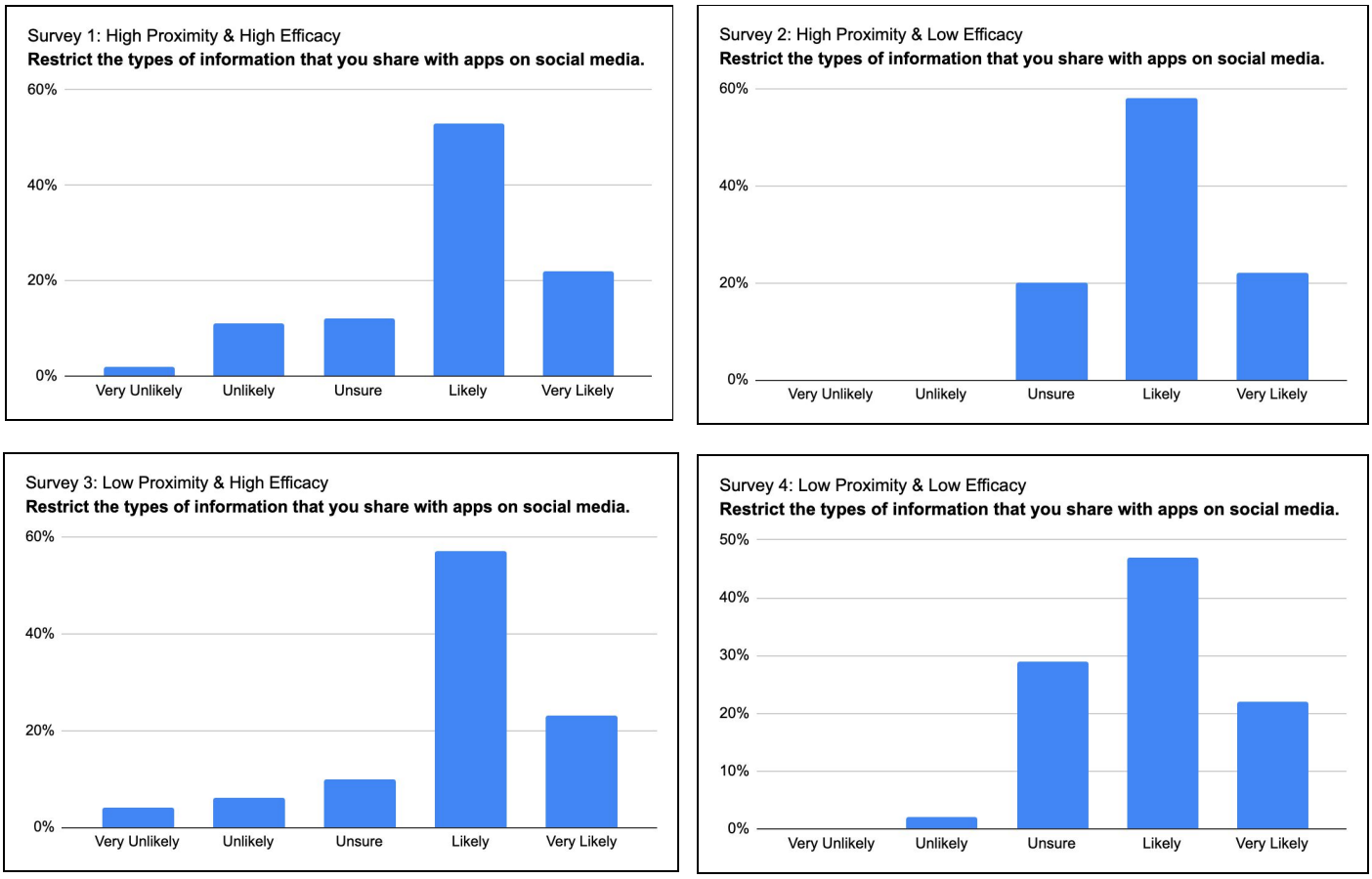
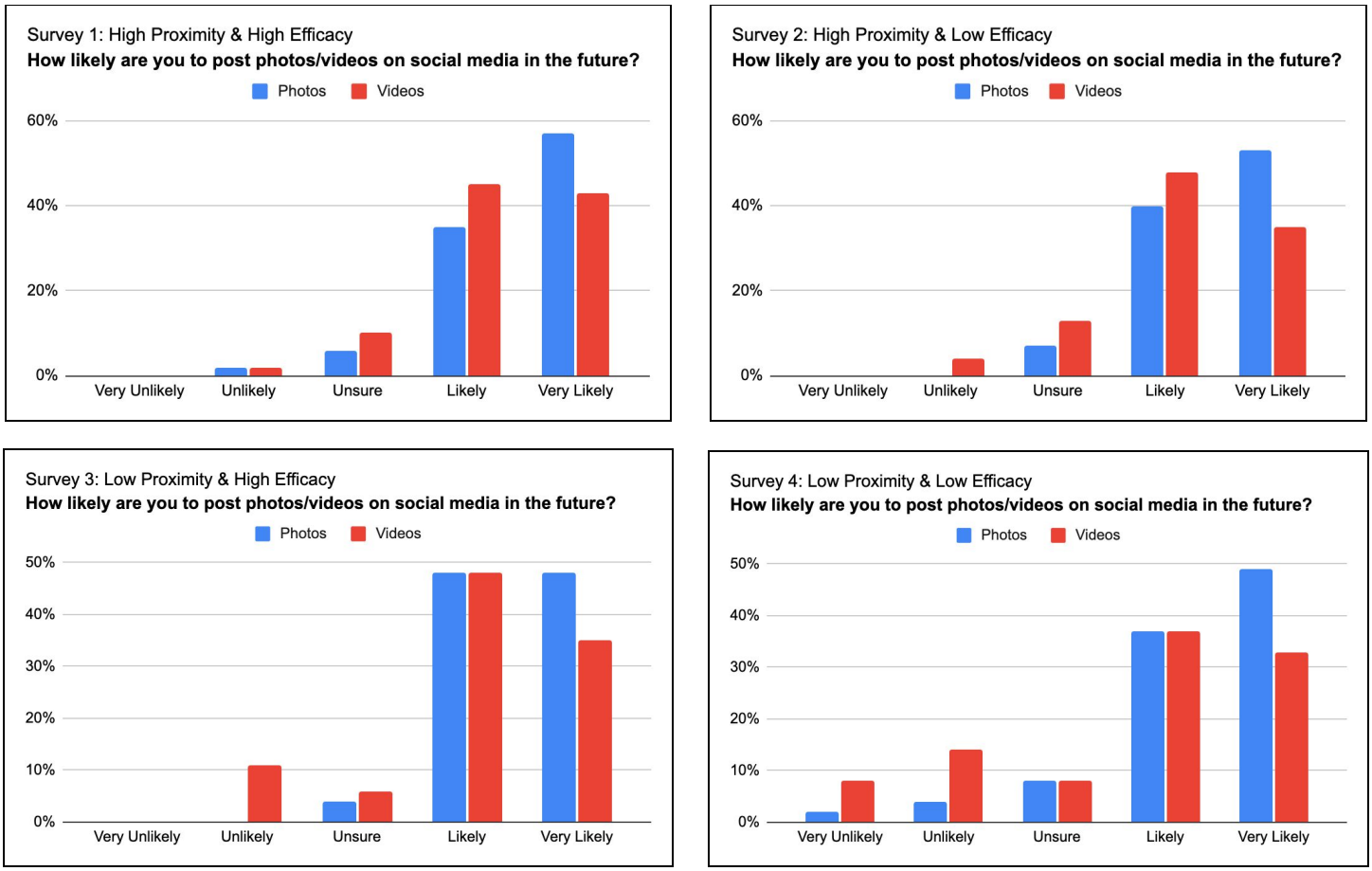

Survey 1: High Proximity & High Efficacy
Restrict the types of information that you share with apps on social media.
60%
40%
20%
0%
Very Unlikely
Unlikely
Unsure
Likely
Very Likely
Survey 2: High Proximity & Low Efficacy
Restrict the types of information that you share with apps on social media.
60%
40%
20%
0%
Very Unlikely
Unlikely
Unsure
Likely
Very Likely
Survey 3: Low Proximity & High Efficacy
Restrict the types of information that you share with apps on social media.
60%
40%
20%
0%
Very Unlikely
Unlikely
Unsure
Likely
Very Likely
Survey 4: Low Proximity & Low Efficacy
Restrict the types of information that you share with apps on social media.
50%
40%
30%
20%
10%
0%
Very Unlikely
Unlikely
Unsure
Likely
Very Likely

Figure 4:


Survey 1: High Proximity & High Efficacy
How likely are you to post photos/videos on social media in the future?
Photos
Videos
60%
40%
20%
0%
Very Unlikely
Unlikely
Unsure
Likely
Very Likely
Survey 2: High Proximity & Low Efficacy
How likely are you to post photos/videos on social media in the future?
Photos
Videos
60%
40%
20%
0%
Very Unlikely
Unlikely
Unsure
Likely
Very Likely
Survey 3: Low Proximity & High Efficacy
How likely are you to post photos/videos on social media in the future?
Photos
Videos
50%
40%
30%
20%
10%
0%
Very Unlikely
Unlikely
Unsure
Likely
Very Likely
Survey 4: Low Proximity & Low Efficacy
How likely are you to post photos/videos on social media in the future?
Photos
Videos
50%
40%
30%
20%
10%
0%
Very Unlikely
Unlikely
Unsure
Likely
Very Likely

Discussion

This study presented new information about the role that artificial intelligence plays in social media and privacy protection. Throughout the four questionnaires, the majority of each group felt that social media platforms such as Instagram and Facebook, are unreliable for protecting personal information. The majority of respondents had more followers than they were following. This means that they could be unaware of who their extra followers are. Many participants had 900+ followers and this presents many risks on social media. On social media, the more followers a person has, the more likes they will get, and the more dopamine will be released to the brain. The human desire to experience instant gratification has led to social media users to become careless with their privacy. Social media provides no protection from people stealing your photos and users' information is constantly being sold to third party advertisers. Users' personal information on social media is not protected. Self-disclosure on social media allows people to fulfill their social need for connection with others. Connecting with people is beneficial but social media makes users vulnerable and presents risks. Deepfakes provide an opportunity for people with malicious intentions to use other's pictures to spread misinformation and cause chaos.

Respondents were asked to rate their awareness of the program Deepfakes and the majority had never even heard of this artificial intelligence development. All of the participants were college educated at a private institution. If college educated students were unaware of this new artificial intelligence program, then it can be assumed that the general public is also unaware of this new threat. It is essential that the public is educated about the dangers of social media privacy, Deepfakes, and steps they can take to protect their personal information online. The

majority of participants did not trust Instagram or Facebook to protect their information from unauthorized use. This demonstrates how college students are aware that social media presents risks but they are still willing to post their information online. The results show that the majority of respondents feel that the benefits of social media outweigh the risks. As the Deepfakes program continues to develop and becomes easier to use, it'll be essential for social media users to understand the dangers of posting personal information. Even though social media users are aware that their information can be used and manipulated, many people have the mentality that it will not happen to them.

The four conditions had varying levels of proximity and efficacy regarding the threat of Deepfakes: condition one - high proximity/high efficacy, condition two - high proximity/low efficacy, condition three - low proximity/high efficacy, & condition four - low proximity/low efficacy. The proximity was measured using the construal level theory, which involves the psychological distance of a threat. Prior research used this theory to study climate change and long term disasters. This study uses this theory to examine how differently college students view a privacy risk when the psychological distance is altered.

Conditions one and three both involved high efficacy, meaning they had more self-control over the threat. Participants were told if they limit their pictures on social media and keep their accounts private, they can avoid becoming a victim of Deepfakes. This resulted in respondents being more unsure about the threat compared to condition two and four that had a low efficacy. In conditions two and four, they were told that there were no effective ways to escape Deepfakes and that information on the Internet remains in the system forever. This resulted in respondents being more likely to restrict the types of information they share on social media in the future. Across the conditions, many respondents viewed Deepfakes as a real

problem. Condition three had the lowest amount of participants say it was a real problem. This is because the conditions had low proximity and high efficacy. Based on the given condition, these respondents view Deepfakes as a distant threat and an issue they can have some control over.

Respondents were asked about future behavior changes they would make regarding the information they post on their social media pages. The social cognitive theory explains how people are more likely to make behavioral change when there is social support and accurate information provided. Throughout the conditions, respondents plan to unfollow and unfriend certain people on their social pages, as well as, restrict the types of information they post online. There was a mixed response from conditions when asked if they would untag themselves from their friends social media posts in the future. A large percentage of respondents in condition one disagreed, condition two had an equal amount agree and disagree, condition three had many agree and condition four was mostly unsure. It was interesting to see how different the questionnaire results were for this question due to the variations of proximity and efficacy. Condition one and two both had high proximity, meaning that the threat of Deepfakes was close to them. It was surprising to see that respondents in these conditions disagreed to untag themselves in others posts. The respondents in condition four with low proximity and low efficacy were unsure about whether to untag themselves in future posts. Even though Deepfakes were viewed as a distant threat, this group still had no control over the situation (low efficacy), which led them to be unsure. It was also interesting to see how many in condition three were likely to untag themselves, when they had the condition that had low proximity and high control.

The majority of respondents across the four conditions said that they plan to delete some information from their current social media. In addition, almost all respondents said they were likely to continue posting pictures and videos on their social media. Almost half of all

respondents agreed that it is easy for them to reduce online privacy risks, and the other half was unsure. According to the construal level theory, people are more likely to change their behavior when the threat is closer. The results demonstrate that almost all respondents will continue to post photos and actively use social media. Many respondents plan to make their accounts private and limit the people they friend/follow. Most of the respondents have never heard of the program, Deepfakes, so it makes sense as to why they still do not view it as a close threat to them.

Limitations and Suggestions for the Future

This online experiment presented some limitations that can be improved in future studies. In order to receive more accurate results involving proximity and efficacy a large sample size would be necessary. It was evident that proximity and efficacy influenced the results but the differences would be clearer with more respondents. All of the respondents were ages 18-22 and college educated. It would be beneficial to see how people are different ages and different education levels would view the threat of Deepfakes. For example, Baby Boomers may not use social media as often as Millennials and GenZ. This may result in Baby Boomers not being as concerned or maybe they would solely be concerned for their children's privacy. Future research should also focus on including different ethnicities, incomes, and genders. This can be accomplished through randomized sampling. In this study, it was primarily white female students and many respondents had a household income of 100,000+. Convenience sampling and systematic sampling was used in this study to achieve the most results in a limited amount of time. Of the respondents 170 were randomized and the remaining were chosen through

convenience. To provide better results it would be beneficial to randomize all the participants using systematic sampling.

Conclusion

Deepfakes are on the rise with no regulation or legal protection. Now that more people are aware of this new AI program, how will they now manage their social media account with all of this new information? Respondents already did not trust social media companies such as Facebook and Instagram to protect their information online. It's interesting that social media users are aware of the privacy threats, and still continue to actively post information to these sites. The desire for personal connection with others causes people to overlook the risks involved with posting information on social media.

Proximity had a limited impact while efficacy had a noticeable influence on respondents' results. The lack of control is what concerned participants more than the distance of the threat. Respondents should be concerned about the lack of control they have regarding Internet privacy. In China, there is already an app called Zao that uses Deepfakes programming. It only requires a few selfies with different facial expressions to put your face in a well-known movie. This demonstrates how Deepfakes are already advancing and it will soon become difficult to identify which videos are inauthentic. Judges and juries will need to be educated and have a substantial amount of evidence to determine which videos are fake. Deepfakes are an issue that everyone needs to be educated and informed about.

-

References

Bazarova, N. N., & Choi, Y. H. (2014). Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication, 64*(4), 635-657. doi:10.1111/jcom.12106

Brügger, A., Morton, T. A., & Dessai, S. (2016). "Proximising" climate change reconsidered: A construal level theory perspective. *Journal of Environmental Psychology, 46*, 125-142. doi:10.1016/j.jenvp.2016.04.004

Cho, J., Chan, K., & Adali, S. (2015). A survey on trust modeling. *ACM Computing Surveys, 48*(2), 1-40. doi:10.1145/2815595

Fallis, D. (2020). The epistemic threat of deepfakes. *Philosophy & Technology*. 24(1), 1-21. doi:10.1007/s13347-020-00419-2

Fiedler, K. (2007). Construal Level Theory as an Integrative Framework for Behavioral Decision-Making Research and Consumer Psychology. *Journal of Consumer Psychology, 17*(2), 101-106. Retrieved October 21, 2020, from http://www.jstor.org/stable/27609638

Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S., & Balissa, A. (2017). Privacy by designers: Software developers' privacy mindset. *Empirical Software Engineering, 23*(1), 259-289. doi:10.1007/s10664-017-9517-1

Henderson, M. D., Wakslak, C. J., Fujita, K., & Rohrbach, J. (2011). Construal level theory and spatial distance: Implications for mental representation, judgment, and behavior. *Social Psychology, 42*(3), 165-173. doi:http://dx.doi.org/10.1027/1864-9335/a000060

Kietzmann, J., Lee, L. W., Mccarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat?

Business Horizons, 63(2), 135-146. doi:10.1016/j.bushor.2019.11.006

Kumar, S. N., Saravanakumar, K., & Deepa, K. (2016).

On privacy and security in social media – A comprehensive study. *Procedia Computer Science*, 78, 114-119. https://doi.org/10.1016/j.procs.2016.02.019.

Liberman, N., Trope, Y., & Wakslak, C. (2007). Construal Level Theory and Consumer Behavior. *Journal of Consumer Psychology, 17*(2), 113-117. Retrieved October 21, 2020, from http://www.jstor.org/stable/27609640

Lin, H., & Chang, C. (2018). What motivates health information exchange in social media? The roles of the social cognitive theory and perceived interactivity. *Information & Management, 55*(6), 771-780. doi:10.1016/j.im.2018.03.006

Lwin, M. O., Li, B., & Ang, R. P. (2012). Stop bugging me: An examination of adolescents' protection behavior against online harassment. *Journal of Adolescence*, 35(1). 31-41. https://doi.org/10.1016/j.adolescence.2011.06.007.

Marwick, A. E., & Boyd, D. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, 16(7), 1051–1067. https://doi.org/10.1177/1461444814543995

Ratten, V., & Ratten, H. (2007). Social cognitive theory in technological innovations. *European Journal of Innovation Management, 10*(1), 90-108. doi:http://dx.doi.org/10.1108/14601060710720564

Riley, W. T., Martin, C. A., Rivera, D. E., Hekler, E. B., Adams, M. A., Buman, M. P., Pavel, M.,

& King, A. C. (2016). Development of a dynamic computational model of social cognitive theory. *Translational behavioral medicine*, *6*(4), 483–495. https://doi.org/10.1007/s13142-015-0356-6

Turcotte, J., York, C., Irving, J., Scholl, R. M., & Pingree, R. J. (2015). News recommendations from social media opinion leaders: Effects on media trust and information seeking. *Journal of Computer-Mediated Communication, 20*(5), 520-535. doi:10.1111/jcc4.12127

Uszynski, M. K., Casey, B., Hayes, S., Gallagher, S., Purtill, H., Motl, R. W., & Coote, S. (2018).

Social cognitive theory correlates physical activity in inactive adults with multiple sclerosis. *International journal of MS care*, *20*(3), 129–135. https://doi.org/10.7224/1537-2073.2016-111

Venema, A. E., & Geradts, Z. J., PhD. (2020). Digital forensics, deepfakes, and the legal process. *Scitech Lawyer, 16*(4), 14-17,23. Retrieved from https://marist.idm.oclc.org/login?url=https://www.proquest.com/docview/2439640810?accountid=28549

Vogel, D. L., & Wester, S. R. (2003). To seek help or not to seek help: The risks of self-disclosure. *Journal of Counseling Psychology, 50*(3), 351-361. doi:http://dx.doi.org/10.1037/0022-0167.50.3.351

Yadlin-Segal, A., & Oppenheim, Y. (2020). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence: The International Journal of Research into New Media Technologies, 5*(1), 1-16. doi:10.1177/1354856520923963

Appendix A: Four Conditions

Condition 1: High Proximity & High Efficacy

Deepfakes is an artificial intelligence system that can create fake images and videos that look completely real. Anyone can upload images to this program to create a fake digital person. It is a threat to everyone's privacy, and no one is safe. Celebrities' faces are being put onto porn stars' bodies to create fake videos. Your friends can put your face onto any video and spread false information. Even though Deepfakes are dangerous, they can be completely prevented. Limiting your pictures on social media and keeping your accounts private can protect you from this threat.

Condition 2: High Proximity & Low Efficacy

Deepfakes is an artificial intelligence system that can create fake images and videos that look completely real. Anyone can upload images to this program to create a fake digital person. It is a threat to everyone's privacy, and no one is safe. Celebrities' faces are being put onto porn stars' bodies to create fake videos. Your friends can put your face onto any video and spread false information. What is even worse, is that there are no effective ways to prevent this threat on the Internet. Information that is posted on the Internet cannot be removed and will remain in the system forever.

Condition 3: Low Proximity & High Efficacy

Deepfakes is a new artificial intelligence system that is being used to create comedy videos of celebrities. A Deepfake video was created with Steve Buscemi's face on Jennifer Lawrence's body while she talked about her favorite housewives. For now, Deepfakes are being used for entertainment purposes. Even though Deepfakes are dangerous, they can be completely prevented. Limiting your pictures on social media and keeping your accounts private can protect you from this threat.

Condition 4: Low Proximity & Low Efficacy

Deepfakes is a new artificial intelligence system that is being used to create comedy videos of celebrities. A Deepfake video was created with Steve Buscemi's face on Jennifer Lawrence's body while she talked about her favorite housewives. For now, Deepfakes are being used for entertainment purposes. What is even worse, is that there are no effective ways to prevent this threat on the Internet. Information that is posted on the Internet cannot be removed and will remain in the system forever.

Appendix B: Questionnaire

How Often do you use the following social media platforms? Facebook, Instagram, Twitter, Youtube, add the popular social media among college students *1= Never 2 = Several times a year 3 = Several times a month 4 = Several times a week 5 = Several times a day*

Social media experience:

How many friends do you have on Facebook?

How many people do you follow on Instagram?

How many people follow you on Instagram?

Knowledge about Deepfakes:

How much do you know about Deepfakes?

1 = I have never heard of it. 2= I heard of it, but know little. 3 = I know some of it. 4 = I know pretty much 5 = I am very familiar with it.

Privacy Concern

How much do you agree with the following statements? 1 = Strongly disagree; 5 = Strongly agree

I feel uncomfortable when information is shared with advertisers or third parties without permission.

I am concerned about misuse of personal information.

I feel fear that information may not be safe while stored.

I believe that personal information is often misused.

I think companies share information without permission

<u>Please read the statement below and then answer the following questions.</u>
Deepfakes is an artificial intelligence system that can create fake images and videos that look completely real. Anyone can upload images to this program to create a fake digital person. It is a threat to everyone's privacy, and no one is safe. Your friends can put your face onto any video and spread false information. Even though Deepfakes are dangerous, it can be completely

prevented. Limiting your pictures on social media and keeping your accounts private can protect you from this threat

Perceived susceptibility: *1 = Very unlikely 5 = Very likely*

How likely do you think these issues will happen to you?
Receiving hate emails
Being threatened online
Receiving unpleasant sexual remarks online
Someone pretending to be me online
Someone publishing my personal information online with bad intentions
Someone posting my personal photos/videos online with the intention to harm me
Someone posting negative rumors or inflammatory remarks about me online

Trust:

How much do you agree with the following statements? 1= strongly disagree 5=strongly agree

Social media such as Facebook and Instagram is trustworthy;

I can count on social media companies such as Facebook or Instagram to protect customers' personal information from unauthorized use;

social media companies such as Facebook or Instagram can be relied on to keep its promises.

Psychological distance:

To me, threat caused by Deepfakes feels very close … very distant

Skepticism: *1= strongly disagree 5=strongly agree*

Experts believe that Deepfakes is a real problem

The media is too alarmist about issues to do with the threat caused by Deepfakes

The evidence for threat caused by Deepfakes is unreliable

I am uncertain if the risks caused by Deepfakes is happening to me

I do not believe risks caused by Deepfakes is a real problem

Personal Efficacy: *(1=strongly disagree; 5=strongly agree)*

I am able to effectively protect my privacy on the Internet.

Taking actions to reduce online privacy risks is easy for me

I can easily adopt the measures to protect my online privacy.

General action efficacy beliefs: *(1=strongly disagree; 5 =strongly agree)*

If everyone does their bit, we can reduce privacy risks on the Internet.

Individual behavior change (e.g., don't share pictures on the Internet) is effective in reducing privacy risks in the society.

Introducing new online privacy protection methods will significantly decrease online privacy risks.

Behavioral Intentions:

How much do you agree with the following statements? *1= strongly disagree 5=strongly agree*

I plan to limit access to my personal information on social media so that not everyone can view it.
I plan to NOT reply to messages from strangers on the Internet.
I plan NOT to disclose my personal data on public websites.

Likelihood of Self Disclosure:

How likely will you post the following information on social media in the future? *(1 = Very unlikely; 5 = Very likely)*

Home address, photos, videos; relationship status, real name, email address, cell phone number, Instagram user ID, religion, personal interest or hobbies such as movies, TV shows, music, books.

Privacy protection behavior:

How likely will you have the following behavior in the future? *(1 = Very unlikely; 5 = Very likely)*

Untag photos/videos of you on social media.

Delete information from your social media.

Unfriend or unfollow people on social media

Restrict access to your social media accounts so that only your friends can see them

Make your status updates private and allow only your friends to see them

Remove an app from your social media account because of the information it collects about you.

Restrict the types of information that you share with apps on social media

Install an additional application to prevent third parties on social media from tracking you and displaying targeted advertisements.

Demographics:

Gender

Age

Race

Ethnicity

House income