

K-pop 아이돌의 딥페이크 포르노그래피에 대한 사람들의 부정적인 감정 연구

왕 소영*, 김 성철^o

How Do People Feel about Deepfake Videos of K-Pop Idols?

Soyoung Wang*, Seongcheol Kim^o

요 약

딥페이크는 인공지능 기술을 이용하여 원본 이미지나 동영상에 다른 이미지를 합성함으로써 조작된 미디어 콘텐츠를 만드는 것을 의미한다. 딥페이크 기술의 실용적인 가치에도 불구하고 딥페이크가 초래할 수 있는 위협이 크고 다양하므로 딥페이크에 대한 언론과 대중의 관심이 증가하고 있다. 본 논문은 먼저 딥페이크가 무엇인지, 왜 문제인지를 설명한다. 특히, 케이팝 아이돌의 딥페이크 포르노그래피에 연구의 초점을 맞추므로써 딥페이크 포르노그래피에 대한 사람들의 부정적인 감정을 살펴본다. 동영상 이용자 300명을 대상으로 온라인 설문조사를 실시하고 그 결과를 분석한 결과, 감정 이입적 관심, 포르노그래피에 대한 잦은 노출, 그리고 성별이 딥페이크 포르노그래피에 대한 사람들의 분노와 죄책감에 영향을 미친다는 것을 알 수 있었다. 또, 성희롱 태도, 케이팝 아이돌에 대한 관여도, 포르노그래피 이용이 성별마다 차이를 파악할 수 있었다. 딥페이크에 대한 사람들의 부정적인 감정에 영향을 미치는 요인을 실증적으로 규명함으로써 본 연구는 딥페이크 문제를 해결하기 위해서 정부, 산업, 사회의 전방위적인 대처가 중요함을 시사하고 있다.

키워드 : 딥페이크, 딥페이크 포르노, K-pop 아이돌, 부정적인 감정

Key Words : Deepfake, Deepfake pornography, K-pop idols, Negative emotions

ABSTRACT

Deepfake refers to the comprehensive process of creating hyper-realistic manipulated media content using deep learning and artificial intelligence(AI). Despite the practical value behind deepfake technology, significant attention is paid to how deepfakes are a tremendous threat to not only media figures but also the general public. The technology's ability to generate duplicates of real faces and create new authentic images raises concerns over privacy, identity, and ethical issues. This study is designed to provide an explanation of deepfakes and the current misuses of deepfake technology. By specifically focusing on deepfake pornography of K-pop idols, the study aims to explore how people feel about fabricated content. The results of the study demonstrate that empathic concern, pornography exposure, and gender are crucial predictors that influence viewers' emotions of anger and guilt toward K-pop idols' deepfake pornography videos. Substantial gender differences are found in pornography usage, celebrity involvement, and attitude toward sexual harassment. As one of the early studies to empirically investigate people's negative feelings about deepfake content, this study suggests the need for collaborative efforts from the society, industry, and content consumers to effectively resolve the deepfake pornography issue.

※ 본 연구는 대한민국 교육부와 한국연구재단의 지원(NRF-2019S1A3A2099973)과 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT연구센터지원사업의 연구결과로 수행되었음(IITP-2020-0-01749).

• First author : School of Media and Communication, Korea University, soyoung1007@korea.ac.kr, 학생회원

^o Corresponding author : School of Media and Communication, Korea University, hiddentrees@korea.ac.kr, 종신회원

논문번호 : 202110-268-0-SE, Received August 30, 2021; Revised November 19, 2021; Accepted November 23, 2021

I. Introduction

Recent advancements in generative models, such as generative adversarial network (GAN) and variational autoencoder (VAE) of artificial intelligence (AI), has enabled people to fabricate videos and, in extreme cases, manipulate media. With just a laptop, a mobile application, and appropriate software, anyone can merge different images and audio into visual content. This process of using deep learning and AI to create hyper-realistic manipulated content by combining, replacing, and superimposing various forms of media is called deepfake^[1,2].

Deepfake technology has great potential in the entertainment industry, especially for special effects and face editing in the post-production process. For example, the technology is used to de-age an actor as in 'The Irishman'(2019) where Robert De Niro looks thirty years younger than his actual age^[3]. While there are valuable opportunities behind deepfake technology, several incidents prove that deepfakes are undoubtedly a tremendous threat to media figures and the general public. For example, deepfake videos are used by politicians to spread fake news and manipulate public opinion, whereas deepfake voices are used by hackers to commit financial fraud crimes^[4]. Such manipulation using deepfake is causing damage to individual and organizational images, and is a severe threat to users who consume various videos online. Furthermore, deepfake pornographic videos (DPVs), which synthesize celebrities' faces onto pornographic actors' bodies without consent is regarded as a critical drawback of the deepfake technology.

Recent studies on deepfakes primarily concentrate on the technology perspective of deepfake creation or detection^[5,6]. Legal analysis of deepfake incidents are also seen in recent studies^[7,8]. While there are studies examining specific type of deepfakes, such as TweepFake^[9], deepfake news^[10], or deepfake pornography^[1,11], there are insufficient studies focusing on Internet users' understanding of deepfakes.

Therefore, this study aims to explore users'

emotions toward deepfake content by explicitly focusing on the factors that yield negative emotional responses toward K-pop idols' deepfake pornography. K-pop idols were chosen as the main subjects of this study as 96 percent of deepfake videos are DPVs, of which 25 percent target female K-pop idols^[4]. The impact of DPVs is assumed to be detrimental to Korean female idols as well as Korean Internet users that have the possibility of being exposed to DPVs while consuming videos for information or entertainment^[2,4,12,13]. In a situation where negative social phenomena caused by deepfake is more prominent than the technology's rosy future, it is important to examine Internet users' perceptions toward controversial technology. With the study's findings, future researchers and related stakeholders may wisely deal with the innovative technology.

This study conducted an online survey to examine the predictors that arouse negative emotions toward K-pop idols' deepfake pornography. Negative emotions are anticipated to be influenced by people's indigenous attributes created by their attitudes, Internet user behavior, and general socio-demographic characteristics. Based on previous literature, sexual harassment attitude, celebrity involvement, empathetic concern, video exposure, pornography exposure, age, gender, and education level were selected as the key predictors to create negative emotions toward K-pop idol DPVs. This study further examines the different emotional responses of men and women to pornographic deepfake content, along with sexual harassment attitude, celebrity involvement, and pornography usage.

This study is constructed as follows. Following the introduction, Section 2 will explain deepfakes, current issues, and why fabricated sensual contents matter. Then, Section 3 describes the methodology, and Section 4 shows how people feel about deepfake pornography. Implications and conclusion of the study will be presented in Section 5.

II. Literature Review

2.1 Deepfakes and the possible threats

Deepfake is a portmanteau of “deep learning” and “fake,” and refers to the outcome of AI that transposes one’s face onto the body of another to create hyper-realistic fake videos^[1,2,14]. Hence, deepfakes are regarded as deceptive media that can mislead the public^[14].

Deepfakes use generative adversarial networks (GANs), where two artificial networks simultaneously and competitively train with each other^[15], for image-to-image translation. Using the same dataset of images or videos, the generator aims to generate realistic data that is indistinguishable from the original content; whereas the discriminator tries to discover the manipulated data that the generator has produced^[5,15]. This training process becomes highly sophisticated when immense amount of high-quality data is put in repeatedly. Insertion of images with various facial expressions is thus anticipated to train the deepfake technology to produce more authentic replica images of individuals^[11,17].

The technology’s potential was discouraged when a user in Reddit utilized open deepfake algorithms to synthesize the faces of famous female actors into pornography^[1]. The technology’s misuse is also seen in political and economic sectors. For instance, a deepfake video of President Obama disparaging Donald Trump in 2018 raised concern about how anyone can be a victim of manipulated content^[4]. In addition, a coup was made in Gabon after the government released a deepfake video of President Ali Bongo’s New Year’s address^[4]. The surprise appearance of the president, who had been absent due to health issues, prompted speculation from political opponents and sparked public outrage towards the government. Increasing number of deepfake audio technology abuse cases is also observed at the individual and organizational level, such as financial fraud scandals^[4]. The evolution of deepfakes and the subsequent emergence of hyper-realistic fake content is becoming a danger to global society.

2.2 Deepfake pornographic videos (DPVs)

DPVs are non-consensual and sexually humiliating fake videos that transplant the faces of celebrities onto the bodies of pornography actors^[1,11,16]. DPVs are one of the many ways to digitally objectify and demean women in media; non-consensual sexual imagery in online spaces has long been the subject of discussion^[11]. Some scholars argue that DPVs reinforce the idea of women existing as sexual objects^[11,17,18]. The previous study discovered that the demand for K-pop female idols in DPVs derives from the secret desires of fervent fans who see idols as their sexual partners^[17]. Kim reveals that the sexual objectification of K-pop idols satisfies the hidden desires of men by perpetuating and enforcing the images of subordinated femininity^[19].

The damaging effects of DPVs seem to be expanded by the specificity of cyberspace and deepfake technology. Specifically, the online environment allows DPVs to be distributed rapidly to a wide range. The anonymity within online spaces motivates DPV creators to target K-pop idols and media figures without severe guilt or sensitivity to digital crime issues. Additionally, deepfake technology exacerbates the problem because any individual can become a victim in DPVs. DPVs can be produced only with a sufficient corpus of the target’s facial images, as sophisticated source codes, AI algorithms, or related software are freely accessed online.

2.3 Negative Emotions

Just as victims of DPVs are anticipated to suffer great emotional pains, Internet users who face them in online spaces would be likely to evoke strong negative emotions toward sexually abusive deepfake content made without consent. The extant studies have discovered that the concept of negative moral emotions can reveal some reasons why users may experience these feelings.

Negative moral emotions are related to individuals’ expression of negative feelings toward the situation with moral violations by compelling perpetrators^[20-22]. These emotions can be divided

into four categories, including other condemning emotions (i.e., anger), self-conscious emotions (i.e., guilt), other suffering emotions (i.e., empathy), and other praising emotions (i.e., elevation, gratitude). In DPV's context, this study judges that the Internet users may feel other condemning (anger) and self-conscious emotions (guilt) to deepfake pornography, a clear example of moral violations related to the victim's agony.

To begin with, the subject of moral anger will head others, particularly the perpetrators who violated moral standards. Since anger is the emotion that individuals experience when insulted, belittled, or hurt by others' actions, Internet users who get frustrated by ill-intentioned fake content may feel anger^[20]. Further, Internet users may feel guilt if they evaluate their watching of DPVs as violating moral rules^[23]. Guilt is likely to be induced when individuals concentrate on their morally repulsive behavior, not themselves, believing that their behavior may bring about a severe threat to the victim^[24].

Based on the previous literature, this study will explore the negative moral emotions of anger and guilt in the context of DPVs.

2.4 Personal experiences and traits

Previous studies have described how individual differences can affect moral sentiments^[25-27]. Scholars have widely agreed that perceptions of new information technologies (IT) may differ according to individuals' experiences and characteristics^[28,29]. Likewise, Internet users' personal experiences and traits would influence emotions after witnessing DPVs online. Considering that DPVs are highly realistic fake content created through new IT, DPVs can elicit negative moral sentiments different from individual characteristics. Accordingly, relevant experiences and characteristics, including previous tolerance for sexual harassment, previous fondness of celebrities, media encounters, empathic concern, and demographics, are expected to impact users' anger and guilt.

Sexual harassment (SH) is a complex concept, but it is widely adopted as behavior that violates or

humiliates an individual with specific sex or gender^[30]. Mazer and Percival proposed that SH is an undesirable and problematic behavior that springs negative consequences for those who experience it^[31]. Some studies have examined potential factors that may affect individual tolerance for SH behaviors. Mazer and Percival focused on gender, discovering men being more tolerant of SH. Tang et al. found the pattern in a more specific context of gender^[32]. The gender equality movement and flexible gender roles had a profound relationship with individuals' tolerance of SH. Thus, individuals' different attitudes toward SH may produce different effects on DPVs that have sexually harassing characters. This study investigates how Internet users with greater SH tolerance can predict fewer emotional responses of anger and guilt to DPVs.

Celebrity involvement (CI) refers to individuals' cognitive, emotional, and behavioral responses to celebrities^[33]. CI is closely connected to media consumption behavior by the public. It is a multidimensional construct, including affinity^[34], parasocial relationship, and identification^[35]. Mainly, affinity is the lowest level of celebrity involvement that stimulates individuals' initial and general liking of celebrities related to their physical attractiveness or credibility^[34,36]. Previous studies have proposed that celebrity involvement is an influential determinant of individuals' emotional and behavioral changes, such as the feeling of symmetric co-emotions and action for celebrities^[35,37]. Thus, this study adopted affinity as a celebrity involvement measurement.

Empathy is the reactions of one person to the perceived experiences of another. Davis argued empathy as a multidimensional construct, including four dimensions under two criteria: cognitive or emotional empathy^[38]. Empathic concern or sympathy for unfortunate others (EC) and personal distress (PD) or pain after seeing others' emotional reactions are measures of emotional empathy. However, past studies have argued that PD is inappropriate to be considered as empathy, thus only included empathic concern in the DPV context^[39,40]. In other words, individuals may feel concerned

about people being maliciously depicted on DPVs and may feel protective towards them^[38].

Many studies have identified that repeated exposure to media, such as online video games or violence, can gradually reduce the responses to an arousal-eliciting stimulus^[41-43]. Repeated viewing of the humorous film^[43], violent media content^[41], and sad film clips related to the death of a loved one^[41], has decreased media users' ability to experience emotions that they felt at first. Some scholars have even shown that individuals who had been regularly exposed to sexually violent films experienced less anxious and depressed feelings for similar content sources^[44]. Thus, this study premises that chronic exposure to emotion-laden media content, including sexually offensive or pornographic content, reduces psychological reactivity.

Demographic factors like gender, age, and education level can influence individuals' emotions toward DPVs. The extant studies have unearthed that gender is a strong predictor of reactions to sexually violent media, and women report intense negative emotions of aversion, guilt, and shame^[45]. Some studies have argued that women's real-life exposure to SH makes them more empathetic to the situations of women being sexually harassed online^{[46][47]}. Furthermore, former literature has explained that individuals' age may also influence their different emotions, as younger generations use the Internet more frequently than adults and the chances of encountering sexually explicit content increases, leading to the possible desensitization of those content^{[48][49][50]}. However, there are also opposite predictions about age. Since the younger generation has faced multiple digital media, they can better specify the content that may infringe on morality. Generation Z, in particular, cited sex crimes online as one of the top three among personal concerns after the pandemic, accounting for 17 percent, which is considerably higher than 9 percent of the millennial generation^[51]. This is expected because the younger generation is often placed in an environment where sex crime prevention systems are weak and spend more time on social media. Additionally, education level is

likely to influence different response levels.

Therefore, in the DPV context, women, more educated, and young people would feel angrier and guiltier toward DPVs.

III. Methodology

3.1 Data Collection

Data for this study were collected by a professional online survey agency, Macromill Embrain, in September 2020. The survey samples were composed of both males and females aged between 18 up to 59. Survey participants were required to have experience in viewing any kind of deepfake video content on an online platform. Before starting the survey, a summarized explanation of the deepfake technology and deepfake pornography was provided at the beginning of the survey questionnaire to give participants a clear understanding of the context of deepfakes. The first part consisted of questions measuring individuals' understanding of deepfake technology, typical and pornographic video watching time and frequency. The second part had a text scenario assuming accidental DPV encounters and relevant screenshots of DPVS. They were provided to measure the subjects' emotional responses to the DPVs. After excluding undependable responses, 293 validated questionnaires were used for the final data analysis. The demographic statistics of the participants are reported in Table 1.

Table 1. Description statistics of the respondents

	Measures	Frequency	Percent
Gender	Male	148	50.5%
	Female	145	49.5%
Age	18-29	76	25.9%
	30-39	64	21.8%
	40-49	76	25.9%
	50-59	77	26.3%
Highest level of education	High School	32	10.9%
	College	33	11.3%
	University	188	64.1%
	Graduate school	40	13.6%
Total		293	100.0%

3.2 Measurement

Sexual Harassment Attitude was measured using the Sexual Harassment Attitude Scale (SHAS) developed by Mazer and Percival and adapted to fit the DPV context^[31]. Celebrity Involvement was the measured level of affinity with media figures, in which the questionnaires were adapted from the work of Wen and Cui. A higher score indicates higher idol involvement^[52]. Empathic concern was measured by Interpersonal Reactivity Index (IRI) developed by Davis, indicating that the higher the score, the higher the empathy^[38]. Negative emotion scales were adapted from the Differential Emotions Scale (DES-IV), widely used to understand multivariate mood states^[53]. The measures of anger and guilt were adapted to fit the context of this study, with the highest score revealing the highest mood state. At first, nine questions consisting of three different sentences of anger, disgust, and guilt were suggested in the survey, yet disgust was excluded from the analysis.

IV. Results

This study performed exploratory factor analysis to verify the validity of variables. Principal component analysis was used, and an orthogonal method (varimax) was adopted to solve the multicollinearity problem and simplify factor loading. It was considered a significant variable at a factor loading of .4 or higher. Table 2 shows the factor analysis results of the variables in which all 11 items had a factor loading value of more than .4. This study computed Cronbach's alphas to assess reliability. As shown in Table 2, most of the values' coefficients exceeded the recommended value of .7, indicating that they were internally consistent.

For the main analysis, the study conducted a multiple-regression analysis to examine which factors impact users' anger and guilt, respectively. The analysis seemed appropriate for this study as it examines both categorical and continuous values. As seen in Table 3, the overall model for predicting determinants of anger showed significance, $F=23.280$, $p<.001$, $R^2 = .396$, R^2 adjusted=.379.

Table 2. Factor analysis and reliability of variables

Variables		Factor loading	CA	
Sexual Harassment Attitude (SHA)	False myths	SHA1	0.800	.826
		SHA2	0.820	
	Problem recognition	SHA3	0.774	.895
		SHA4	0.769	
Celebrity Involvement (CI)	Affinity	CI1	0.899	.922
		CI2	0.904	
		CI3	0.908	
		CI4	0.868	
Empathic Concern (EC)	-	EC1	0.783	.648
		EC2	0.723	
		EC3	0.614	

(Note: CA = Cronbach's alpha)

Table 3. Multiple-regression analysis for anger toward K-pop idol DPVs

Variables		Anger	
		β	t
Attitudes	SHA	-0.096	-1.846 ⁺
	CI	0.044	0.883
	EC	0.285	5.814***
Video Exposure	Porn view	-0.260	-5.063***
	Video view	-0.058	-1.191
General Demographics	Gender	0.333	6.367***
	Age	-0.052	-1.020
	Education level	.000	.008
R Square		0.396	
Adjusted R Square		0.379	
F		23.280	

(Note: ⁺p<.1, *p<.05, **p<.01, ***p<.001)

Among the eight independent variables, empathic concern ($t=5.814$, $p<.001$), porn view ($t=-5.063$, $p<.001$), and gender ($t=6.367$, $p<.001$) were found to arouse respondents' overall anger toward K-pop idols' DPVs. (Table 3) The model for examining emotions of guilt toward DPVs was also found to be significant, $F=16.325$, $p<.001$, $R^2=.315$, R^2 adjusted=.296. The findings showed a similar pattern to anger, in that empathic concern ($t=4.542$, $p<.001$), porn view ($t=-4.306$, $p<.001$), and gender ($t=5.424$, $p<.001$) were critical factors that determine participants' emotions of guilt. Although weakly

supported, sexual harassment attitude was related to both anger ($t=-1.846, p<.1$) and guilt ($t=-1.729, p<.1$). Along with this result, the study found that age ($t=-2.068, p<.05$) significantly affects emotions of guilt (Table 4).

Furthermore, an independent samples t-test was conducted to compare the different emotions of men and women aroused from watching DPVs. A substantial difference between men ($M=3.36, SD=0.91$) and women ($M=4.23, SD=0.82$) were found for anger ($t=-8.61, p<.001$). In addition, men ($M=3.39, SD=0.89$) and women ($M=4.13, SD=0.84$) showed a significant difference for guilt, ($t=-7.35, p<.001$)

A similar test was performed to examine different gender responses for independent variables. For SHA, men ($M=1.90, SD=0.70$) and women ($M=1.55, SD=0.70$) showed a substantial difference ($t=4.17, p<.001$). Differences in men ($M=3.13, SD=0.94$) and women ($M=3.73, SD=0.90$) were also observed for the celebrity involvement factor, ($t=-5.63, p<.001$). For variables related to empathic concern, no substantial dissimilarities were found between men ($M=3.53, SD=0.61$) and women ($M=3.54, SD=0.64$), and video viewing as well, which showed results of men ($M=2.76, SD=1.48$) and women ($M=2.90, SD=1.63$). However, a significant gender difference was revealed in

Table 4. Multiple-regression analysis for guilt toward K-pop idol DPVs

Variables		Guilt	
		β	t
Attitudes	SHA	-0.095	-1.729 ⁺
	CI	-0.005	-0.095
	EC	0.237	4.542***
Video Exposure	Porn view	-0.236	-4.306***
	Video view	-0.042	-0.823
General Demo-graphics	Gender	0.302	5.424***
	Age	-0.112	-2.068*
	Education level	0.060	1.194
R Square		0.315	
Adjusted R Square		0.296	
F		16.325	

(Note: ⁺ $p<.1$, * $p<.05$, ** $p<.01$, *** $p<.001$)

Table 5. Independent sample t-test of different gender responses

Variable	Gender	N	M	SD	t	p
Anger	Men	148	3.36	0.91	-8.61	.000***
	Women	145	4.23	0.82		
Guilt	Men	148	3.39	0.89	-7.35	.000***
	Women	145	4.13	0.84		
SHA	Men	148	1.90	0.70	4.17	.000***
	Women	145	1.55	0.70		
CI	Men	148	3.13	0.94	-5.63	.000***
	Women	145	3.73	0.90		
EC	Men	148	3.53	0.61	-0.15	0.882
	Women	145	3.54	0.64		
porn view	Men	148	2.43	1.58	5.60	.000***
	Women	145	1.54	1.10		
video view	Men	148	2.76	1.48	-0.81	0.420
	Women	145	2.90	1.63		

(Note: ⁺ $p<.1$, * $p<.05$, ** $p<.01$, *** $p<.001$)

pornography viewing behavior between men ($M=2.43, SD=1.58$) and women ($M=1.54, SD=1.10$) ($t=5.60, p<.001$) (Table 5).

V. Discussion and Conclusion

This study focuses on a recent phenomenon that uses deep learning technology to create hyper-realistic manipulated media content. The deepfake technology causes great harm by duplicating images of media figures or even placing them in different body figures. This study specifically explores the use of deepfake technology in producing pornography videos of K-pop idols and examines users' negative emotional responses of anger and guilt. Potential predictors were analyzed using the multiple regression analysis, and different gender responses were examined using the independent samples t-test.

This study found that empathic concern, pornography view, and gender determine the level of negative emotions. Sensitivity toward sexual harassment has been proven to be weakly connected to individuals' anger while age is linked to guilt. Although both men and women express negative

emotions toward DPVs, women exhibit stronger emotions than men.

In detail, given that DPV is one of the examples of sexual harassment online, prior attitude on SH is likely to result in negative feelings of anger and guilt. The findings showed that individuals who were less tolerant toward SH were angrier, resistant, less permissive, and feeling more guilt to online sexual harassment content.

Moreover, the results demonstrated that participants who are more sympathetically concerned and care about others show more negative emotions to DPVs than those with lower empathy. The empathic concern scale itself indicates a feeling of sensitivity or suffering toward unfortunate others who have been hurt, wronged, or harmed. After being exposed to DPVs, intense feelings of anger to malicious perpetrators and guilt for other's misfortune may have been elicited in every case. It was also revealed that less exposure to watching pornography is associated with more negative emotions. This finding is accordant with prior studies; repeated exposure to violent media can gradually lessen negative emotions, known as desensitization to Internet content^{[43][44]}. Presumably, users who watched a considerable number of pornographies repeatedly can be insensitive to the DPVs of K-pop stars. On the contrary, deepfake pornography would be more irritating for those who are scarcely exposed to sexually violent videos, leading to greater intensity of negative arousals.

In terms of demographic factors, more guilt was observed in the younger population regarding the age factor, meaning that a sense of responsibility and alertness is more visible to the digital natives.

Further, gender hugely impacts emotions of anger and guilt, in which women have expressed much stronger negative emotions about DPVs in every context. As most DPVs distributed are deceitfully aimed to humiliate female celebrities, the gender difference is likely to have occurred. Consistent with previous studies, this study supports that women tend to report more negative emotions of anger and guilt in response to porn videos^[45]. At the same time, men tend to seek out or favor pornography

without emotional attachments or a relational concern^[48]. In addition, the results of the independent sample t-test reveal that substantial gender differences are discovered in intolerance toward sexual harassment, celebrity involvement, and pornography exposure. While a women group shows high resistance toward sexual harassment and has a strong connection with a celebrity with K-pop idols, a men group exhibits more experience of viewing pornography. The results pinpoint the gap between men's and women's attitudes on non-consensual pornography and the problem of women objectification in cyberspace.

This study has a few academic implications. As one of the early efforts to focus on users' emotional responses in the DPV context, it created a foothold for future active discussions on Internet users' thoughts, reactions, and behaviors to the dark side of emerging media. DPV is a representative example of technology abuse and can cause severe damage to Internet users. Since images online are collected as primary data to create DPVs, all Internet users who post photos with their faces on social media or personal blogs can become victims of deepfakes. In this respect, it is considered a vital attempt to examine the impact of AI on Internet users and society, which has been overlooked in the academic field.

In addition, this study deals with the concept of moral emotions, especially anger and guilt. This study discovered some individual traits that can predict individuals' emotions. Namely, emotions have played an essential role in providing meaningful explanations for how users may react to the inappropriate media content combined with emerging technology. The media industry is where instantaneous feelings stand out more than the user's perception or knowledge. Moreover, moral emotions are critical in people's reactions to DPVs because they can bring about users' prosocial behaviors in the DPV context. Previous studies have suggested that negative moral emotions are related to individuals' moral decision-making and coping processes in the real world^{[25][27]}. In this regard, individuals' immediate emotional reactions to DPVs

can explain their attitude to media violence, which is expected to result in subsequent actions. Thus, this study's attempt to apply moral emotions in the deepfake technology context enlarged the discussion of human morality, attitudes, and potential behaviors.

This study also provides several practical implications. First, this study advises the need for more educational programs. As digital sex crimes like DPVs have become a severe social problem these days, education on sexual assaults online has become more significant. The findings revealed that users with little tolerance for sexual harassment felt more anger. Users who can react to the perceived experiences of unfortunate others and have had little experience with pornographic videos felt severe anger and guilt when exposed to DPVs. Thus, intensive educational programs related to cyberspace threats, opportunities, and ethics would be critical for internet users to understand emerging media and subsequent products better. In addition, the findings of the independent sample t-test also investigated that responses of men and women are significantly different. Women react more negatively as most of the victims of online sexual violence are female users, and they have more frequent sexual traumatic experiences in their daily lives. Hence, they usually have a better understanding of this situation. In contrast, men would have limited experience and few opportunities to think about digital sex crimes seriously. Therefore, a demonstrative education approach seems to be necessary for men. A proper education would enable internet users to recognize that DPVs are apparent form of digital sex crimes.

At the industry level, entertainment agencies are expected to develop prompt damage prevention measures and proper protection measures for the degraded idols. AI-related firms would pay more attention to the R&D of deepfake detection, identification, or prevention technologies. The government also seems to be responsible for enacting proper regulations and policies and investing in related social institutions. Although the Special Act on the Punishment of Sexual Violence Crimes, a law that strengthened punishment for

deepfake videos, was established on June 25, 2020, it is still difficult to apply it to foreigners. Thus, it seems to be imperative to provide an integrated platform to deal with digital sex crime issues in processes and manuals. Through collaborative efforts of all related stakeholders, including schools, entertainment agencies, AI-based firms, and the government, the harm from DPVs would be alleviated.

This study is not without limitations. As an exploratory study, it suggests several meaningful antecedents of users' emotions. However, other predictors with better theoretical or practical backgrounds can be further included to understand the social phenomena of deepfake technology. In addition, since the sample includes only Korean participants, this study has a limitation in generalizing its findings. Thus further studies might conduct a comparative analysis of viewers from different countries.

References

- [1] C. Gosse and J. Burkell, "Politics and porn: How news media characterizes problems presented by deepfakes," *Critical Stud. Media Commun.*, vol. 37, no. 5, pp. 497-511, Oct. 2020.
- [2] M. H. Maras and A. Alexandrou, "Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos," *The Int. J. Evidence & Proof*, vol. 23, no. 3, pp. 255-262, Oct. 2018.
- [3] J. Aldredge, *Is deepfake technology the future of the film industry?*(2020), Retrieved Jul. 31, 2021, from <https://www.premiumbeat.com/blog/deepfake-technology-future-of-film-industry/>.
- [4] H. Ajder, F. Patrini, and L. Cullen, *The state of deepfakes: Landscape, threats, and impact*, *DeepTrace Labs*(2019), Retrieved Jan. 8, 2020, from https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.
- [5] D. Guera and E. J. Delp, "Deepfake video

- detection using recurrent neural networks,” in *2018 15th IEEE Int. Conf. AVSS*, pp. 1-6, Auckland, New Zealand, Nov. 2018.
- [6] S. Lyu, “Deepfake detection: Current challenges and next steps,” in *2020 IEEE ICMEW*, pp. 1-6, London, United Kingdom, Jul. 2020.
- [7] A. P. Gieseke, “The new weapon of choice: Law’s current inability to properly address deepfake pornography,” *Vanderbilt Law Rev.*, vol. 73, no. 5, pp. 1479-1516, Oct. 2020.
- [8] S. K. Bae, “A study on criminal regulations against deepfake porn,” *The HUFs Law Res. Inst.*, vol. 43, no. 3, pp. 169-187, Aug. 2019.
- [9] T. Fagni, F. Falchi, M. Gambini, A. Martella, and M. Tesconi, “TweepFake: About detecting deepfake tweets,” *Plos One*, vol. 16, no. 5, e0251415, May 2021.
- [10] C. Whyte, “Deepfake news: AI-enabled disinformation as a multi-level public policy challenge,” *J. Cyber Policy*, vol. 5, no. 2, pp. 199-217, May 2020.
- [11] S. Maddocks, “A deepfake porn plot intended to silence me: Exploring continuities between pornographic and ‘political’ deep fakes,” *Porn Stud.*, vol. 7, no. 4, pp. 415-423, Jun. 2020.
- [12] E. Shearer and K. E. Matsa, *News use across social media platforms 2018*(2018), Retrieved Sep. 10, 2020, from <https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>.
- [13] N. Newman, *Digital News Report: South Korea*(2019), Retrieved Sep. 11, 2020, from <http://www.digitalnewsreport.org/survey/2019/south-korea-2019/>.
- [14] B. Paris and J. Donovan, *Deepfakes and cheap fakes*(2019), Retrieved Sep. 12, 2020, from https://apo.org.au/node/259911?utm_source=APO-feed&utm_medium=RSS&utm_campaign=rss-all.
- [15] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, “Generative adversarial networks: An overview,” *IEEE Sign. Process. Mag.*, vol. 35, no. 1, pp. 53-65, Jan. 2018.
- [16] C. Ohman, “Introducing the pervert’s dilemma: A contribution to the critique of deepfake pornography,” *Ethics and Inf. Technol.*, vol. 22, no. 2, pp. 1-8, Jun. 2020.
- [17] E. van der Nagel, “Verifying images: Deepfakes, control, and consent,” *Porn Stud.*, vol. 7, no. 4, pp. 424-429, Mar. 2020.
- [18] E. J. Dickson, *Deepfake Porn Is Still a Threat, Particularly for K-Pop Stars*(2019), Retrieved Nov. 11, 2020, from <https://www.rollingstone.com/culture/culture-news/deepfakes-nonconsensual-porn-study-kpop-895605/>.
- [19] G. Kim, “K-pop female idols as cultural genre of patriarchal neoliberalism: A gendered nature of developmentalism and the structure of feeling/experience in contemporary Korea,” *Telos*, vol. 184, pp. 185-207, Fall 2018.
- [20] C. S. Carver and E. Harmon-Jones, “Anger is an approach-related affect: Evidence and implications,” *Psychological Bulletin*, vol. 135, no. 2, pp. 183-204, Mar. 2009.
- [21] H. B. Lewis, “Shame and guilt in neurosis,” *Psychoanalytic Rev.*, vol. 58, no. 3, pp. 419-438, Fall 1971.
- [22] C. Molho, J. M. Tybur, E. Guler, D. Balliet, and W. Hofmann, “Disgust and anger relate to different aggressive responses to moral violations,” *Psychological Sci.*, vol. 28, no. 5, pp. 609-619, Mar. 2007.
- [23] J. P. Taguey, J. Stuewig, and D. J. Mashek, “Moral emotions and moral behavior,” *Annu. Rev. Psychol.*, vol. 58, pp. 341-372, Jan. 2007.
- [24] R. F. Baumeister, A. M. Stillwell, and T. F. Heatherton, “Guilt: An interpersonal approach,” *Psychological Bulletin*, vol. 115, no. 2, pp. 243-267, Mar. 1994.
- [25] J. Haidt, et al., *The moral emotions*, Handbook of affective sciences, 2003.
- [26] B. Monin, D. A. Pizarro, and J. S. Beer, “Deciding versus reacting: Conceptions of moral judgement and the reason-affect debate,” *Rev. General Psychol.*, vol. 11, no. 2, pp. 99-111, Jun. 2007.
- [27] R. A. Shweder and J. Haidt, “The future of moral psychology: Truth, intuition, and the

- pluralist way,” *Psychol. Sci.*, vol. 4, no. 6, pp. 360-365, Nov. 1993.
- [28] U. Gal and N. Berente, “A social representations perspective on information systems implementation,” *Inf. Technol. & People*, vol. 21, no. 2, pp. 133-154, Jun. 2008.
- [29] H. Xu, “The effects of self-construal and perceived control on privacy concerns,” in *Proc. ICIS 2007*, 125, Montreal, Quebec, Can., Dec. 2007.
- [30] S. H. Lopez, R. Hodson, and V. J. Roscigno, “Power, status, and abuse at work: General and sexual harassment compare,” *The Sociological Quart.*, vol. 50, pp. 3-27, Dec. 2016.
- [31] D. B. Mazer and E. F. Percival, “Ideology or experience? The relationships among perceptions, attitudes, and experiences of sexual harassment in university students,” *Sex Roles*, vol. 20, no. 3, pp. 135-147, Feb. 1989.
- [32] C. S. K. Tang, M. S. Yik, F. M. Cheung, P. K. Choi, and K. C. Au, “How do Chinese college students define sexual harassment?,” *J. Interpersonal Violence*, vol. 10, no. 4, pp. 503-515, Dec. 1995.
- [33] W. J. Brown and M. A. C. de Matviuk, “Sports celebrities and public health: Diego Maradona’s influence on drug use prevention,” *J. Health Commun.*, vol. 15, no. 4, pp. 358-373, Jun. 2010.
- [34] D. C. Giles, “Parasocial interaction: A review of the literature and a model for future research,” *Media Psychol.*, vol. 4, no. 3, pp. 279-305, Nov. 2009.
- [35] W. J. Brown, M. D. Basil, and M. C. Bocarnea, “Social influence of an international celebrity: Responses to the death of Princess Diana,” *J. Commun.*, vol. 53, no. 4, pp. 587-605, Jan. 2006.
- [36] E. Veer, I. Becirovic, and B. A. Martin, “If Kate voted Conservative, would you? The role of celebrity endorsements in political party advertising,” *Eur. J. Marketing*, vol. 44, no. 3/4, pp. 436-450, Apr. 2010.
- [37] W. Wirth and H. Schramm, “Media and emotions,” *Commun. Res. Trends*, vol. 24, no. 3, pp. 3-39, Sep. 2005.
- [38] M. H. Davis, “A multidimensional approach to individual differences in empathy,” *JSAS Catalog Sel. Documents in Psychol.*, vol. 10, pp. 85-104, 1980.
- [39] S. Baron-Cohen and S. Wheelwright, “The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences,” *J. Autism and Develop. Disorders*, vol. 34, no. 2, pp. 163-175, Apr. 2004.
- [40] E. J. Lawrence, P. Shaw, D. Baker, S. Baron-Cohen, and A. S. David, “Measuring empathy: Reliability and validity of the empathy quotient,” *Psychol. Med.*, vol. 34, no. 5, pp. 911-920, Jul. 2004.
- [41] B. Krahé, A. Berger, I. Vanwesenbeeck, G. Bianchi, J. Chliaoutakis, A. Fernández-Fuertes, and A. Zygađło, “Prevalence and correlates of young peoples sexual aggression perpetration and victimisation in 10 European countries: A multi-level analysis,” *Culture, Health & Sexuality*, vol. 17, no. 6, pp. 682-699, Jan. 2015.
- [42] C. A. Anderson and K. E. Dill, “Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life,” *J. Personality and Soc. Psychol.*, vol. 78, no. 4, pp. 772-790, Apr. 2000.
- [43] K. A. Fanti, E. Vanman, C. C. Henrich, and M. N. Avraamides, “Desensitization to media violence over a short period of time,” *Aggressive Behavior: Official J. Int. Soc. for Res. Aggression*, vol. 35, no. 2, pp. 179-187, Jan. 2009.
- [44] D. Linz, E. Donnerstein, and S. M. Adams, “Physiological desensitization and judgments about female victims of violence,” *Human Commun. Res.*, vol. 15, no. 4, pp. 509-522, Jun. 1989.
- [45] H. A. Rupp and K. Wallen, “Sex differences in viewing sexual stimuli: An eye-tracking study in men and women,” *Hormones and Behavior*, vol. 51, no. 4, pp. 524-533, Apr.

2007.

- [46] M. Lindsay, J. M. Booth, J. T. Messing, and J. Thaller, "Experiences of online harassment among emerging adults," *J. Interpersonal Violence*, vol. 31, no. 19, pp. 3174-3195, May 2015.
- [47] N. J. Malovich and J. E. Stake, "Sexual harassment on campus: Individual differences in attitudes and beliefs," *Psychol. Women Quart.*, vol. 14, no. 1, pp. 63-81, Mar. 1990.
- [48] M. Anderson and J. J. Jiang, *Teens, social media & technology 2018*(2018), Retrieved Sep. 13, 2020, from <https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/>.
- [49] E. J. Chan, D. E. McNeil, and R. L. Binder, "Sex offenders in the digital Age," *J. Am. Academy of Psychiatry and the Law*, vol. 44, no. 3, pp. 368-375, Sep. 2016.
- [50] P. M. Greenfield, "Inadvertent exposure to pornography on the Internet: Implication of peer-to-peer file-sharing networks for child development and families," *Applied Develop. Psychol.*, vol. 25, no. 6, pp. 741-750, Nov-Dec. 2004.
- [51] S. H. Oh and S. H. Kim, *Deloitte Global 2021 MZ generation survey*(2021) Retrieved Aug. 16, 2021, from https://www2.deloitte.com/content/dam/Deloitte/kr/Documents/consumer-business/2021/kr_consumer_article_20210706.pdf.
- [52] N. Wen and D. Cui, "Effects of celebrity involvement on young people's political and civic engagement," *Chin. J. Commun.*, vol. 7, no. 4, pp. 409-428, Sep. 2014.
- [53] G. J. Boyle, "Secondary mood-type factors in the Differential Emotional Scale (DES-IV)," *Multivariate Experimental Clinical Res.*, vol. 8, no. 2, pp. 211-220, 1987.

왕 소 영 (Soyoung Wang)



2020년 8월 : 고려대학교 정치
의교학과 졸업
2020년 9월~현재 : 고려대학교
일반대학원 미디어학과 석사
과정 재학
<관심분야> 뉴미디어, 엔터테인
먼트산업, IP비즈니스, 인공지
능 윤리

[ORCID:0000-0002-4154-6494]

김 성 철 (Seongcheol Kim)



1987년 2월 : 서울대학교 경영
학과 졸업
1989년 2월 : 서울대학교 대학
원 경영학석사
1996년 8월 : Michigan State
University Telecommunication
석사

2000년 12월 : Michigan State University Media 박
사

2008년 3월~현재 : 고려대학교 미디어학부 교수

2020년 6월~현재 : Digital Business (Elsevier 발간
국제학술지) Co-Editor

2020년 7월~현재 : 스마트미디어 서비스 연구센터
(SSRC) 센터장, 지능정보기술과 사회문제 연구센
터(CIS) 센터장, 고려대 4단계 BK21 미디어학교
육연구단 단장

<관심분야> 뉴미디어, 디지털비즈니스, 미디어산업

[ORCID:0000-0001-6703-4465]