IDS2935: Privacy in the Digital Age
Matthew Sabag
04/25/25
Effective Word Count: 1954

Blurring the Line Between Real and Fake: Public Perception and Privacy Risks in the Era of AI-Generated Media

Abstract

This critical analysis explores the growing privacy and trust concerns associated with AI-generated media, specifically deepfake videos. As artificial intelligence becomes increasingly capable of creating hyper-realistic audio and visual content, the general public's capacity to identify real from synthetic media is critically being tested. A survey study was implemented to deepen the understanding of the general public's perspectives regarding their privacy and concerns on AI-generated content and their ability to distinguish media samples. Through the statistics of the survey that was conducted, it was discovered that over half of individuals that participated expressed that they had a predetermined concern with AI being used for the spread of misinformation as well as its use in legal and political contexts. It is crucial to emphasize that AI, in the perspective of the public, is capable of malice and may be a threat to the general welfare despite taking any considerable action to set regulations in place. In this paper I will answer the following research question: 'To what extent can AI-generated audio-visual content emulate real-life scenarios in ways that surpass the average human's ability to distinguish between original and artificial media?'

Introduction

We live in an era where media is the dominant form of entertainment. As of today, limited by the technology we have, AI has skyrocketed in popularity. Artificial Intelligence is capable of many things, including ordinary research, casual conversation, tedious tasks, and even audio-visual realistic generated media. In this paper, I will specifically cover the privacy concerns of a type of media generated by these growing AI models: deepfakes. A deepfake is a doctored photo, video, or sound recording made by Artificial Intelligence that resembles a hyper-realistic imitation of an individual or groups of individuals. This unsettling technology is available to anyone with a smartphone, practically anybody nowadays, and can potentially generate synthetic media that is convincing enough to fool the average viewer. This power is exceptionally susceptible to being misused and abused, making me ask myself, “Can people really tell the difference between what is real and what is fake?”

It all started when I was sitting at home, scrolling on my phone as a somewhat-lazy college student typically does, and I came across a video of bodycam footage from the police. Although nothing *interesting* was happening, I still paid no attention to the video itself, not thinking twice about it. When I shifted over to the comments section, I read comments like “If it didn’t say it was AI, I would’ve believed it!” and other similar comments. I thought to myself, “Wait, it's AI?!” In the modern times that we live in, the capabilities of AI are increasing exponentially each day. Today, we see an unsuspecting video of bodycam footage, tomorrow we see bodycam footage of someone committing a crime and spending the rest of their life in jail. The concept of deepfakes really moved me, and encouraged me to conduct a survey that collects data of how perceptive the public is to AI-generated content.

Methods

The method that I used to conduct my study aligns with the format of a survey study. The survey included niche questions that were designed to provide insights into the public's detection accuracy, confidence levels, and attitudes toward the privacy risks associated with AI content. I took advantage of the resources available, one of which is the online survey resource known as Qualtrics. Qualtrics enabled me to create an interactive and comprehensive survey style. Once the survey was completed, I distributed the survey digitally to several friends and family members and asked them to share it with their friends and family. This method gathered traction for my survey rather quickly, with over 150 participants within the first 24 hours of the survey being published. I also established that any adult who consented to participating in the survey was able to take it with no discrimination or bias toward participant types.

The survey was formatted in an organized manner, consisting of three main sections. The first section primarily asked participants if they are over the age of 18 and consented to participating in the survey, followed by about 15 preliminary questions establishing their initial perceptions, concerns, and confidence levels. The second component contained three short clips that asked participants if the footage was real or fake, and how confident they were about their response. Although it was never revealed in the survey, I had many participants reach out to me asking which videos were real and AI. The third and final section was a post-evaluation reflection where participants answered questions about their final opinions and perceptions.

Results

Figure 1

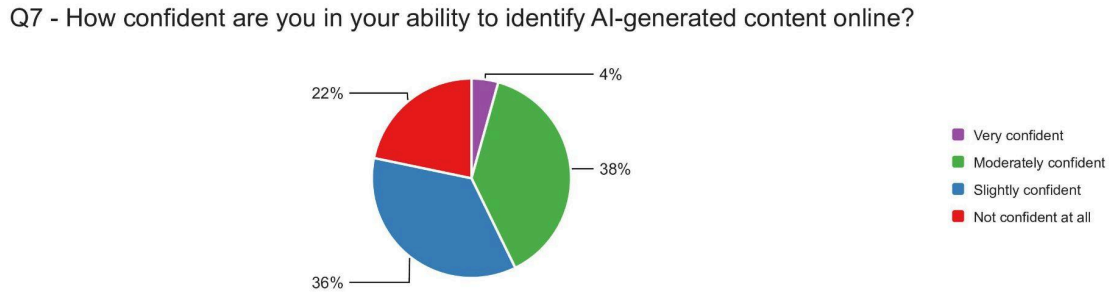


Figure 1 represents how confident participants are in regards to their AI detection skills before viewing the short clips where they are asked to determine if the media is either real or fake.

Figure 2

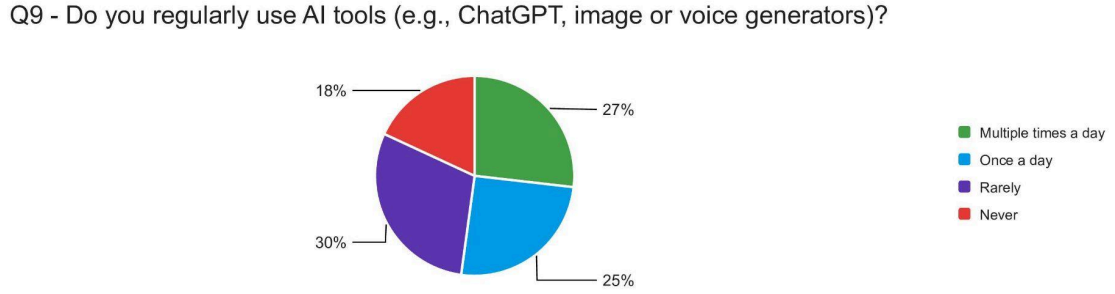


Figure 2 establishes participants' familiarity with Artificial Intelligence and its capability to perform tasks, handle complex directives, and understand a wide variety of material. Based on popularity, it was assumed that participants referred to ChatGPT when asked about their frequent use of AI tools rather than traditional image, voice, and video generators.

Figure 3

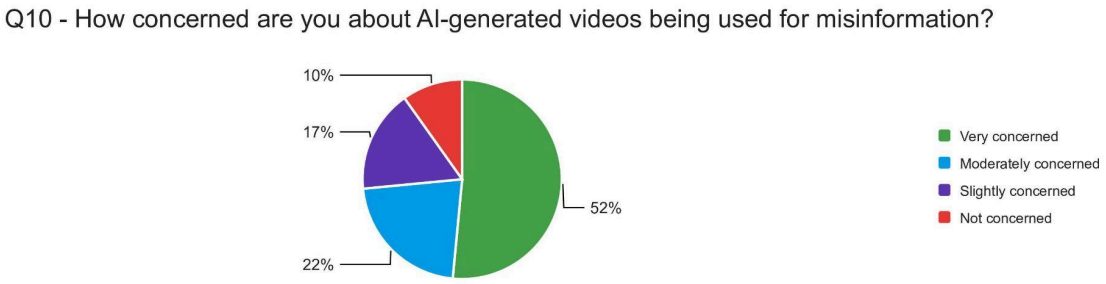


Figure 3 visualizes the apparent concern of AI being misused to spread misinformation and fake news. Only 10% of participants were *not concerned at all*.

Figure 4

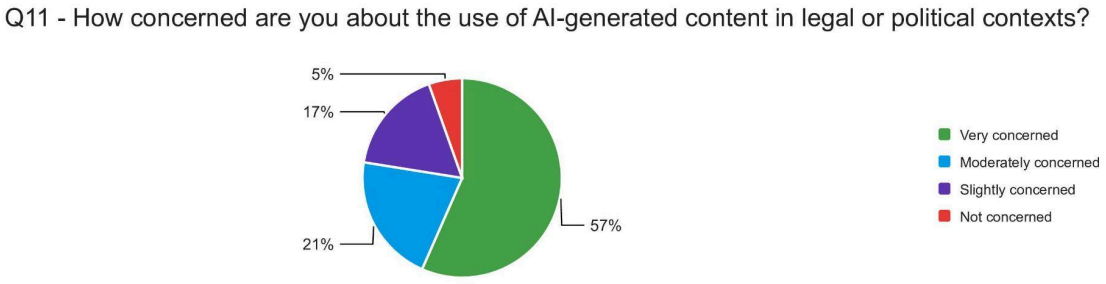
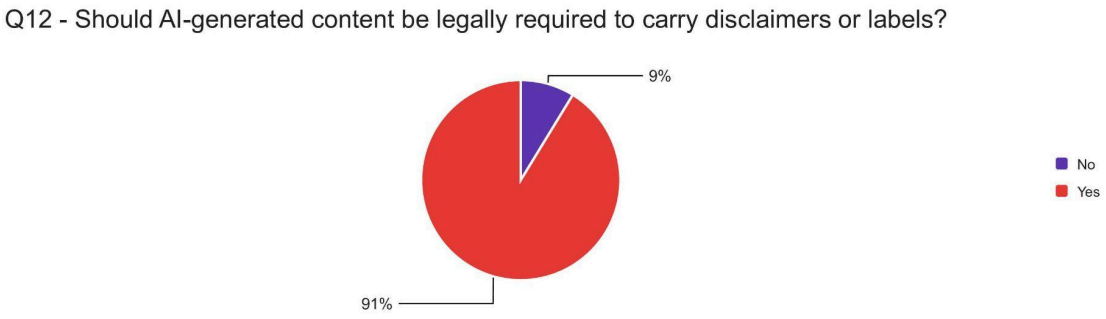


Figure 4 visualizes the apparent concern of AI being used for legal action and political manipulation. Only 5% of participants were *not concerned at all*, half of the previous concern.

Figure 5



The survey response in *Figure 5* indicates whether AI should or shouldn't be legally required to include disclaimers in any posted content or labels when content is generated. Over 90% of participants answered that it should be required by law.

Figure 6

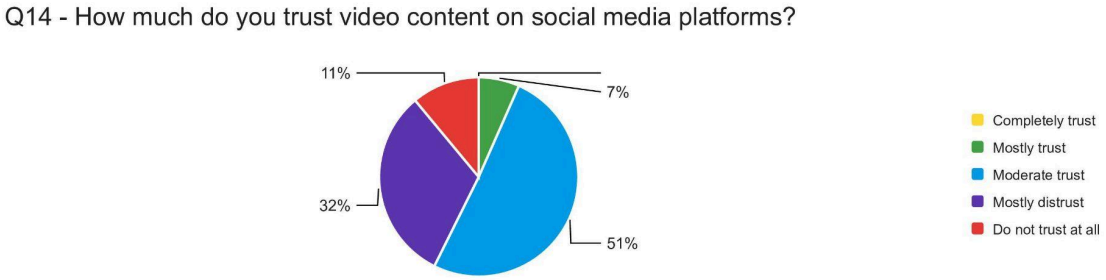


Figure 6 provides a sense for how trustworthy participants find videos on social media platforms to be. 0% of participants completely trust content they find on social media, whereas 58% trust it unlike the remaining 43% that distrust it.

Figure 7

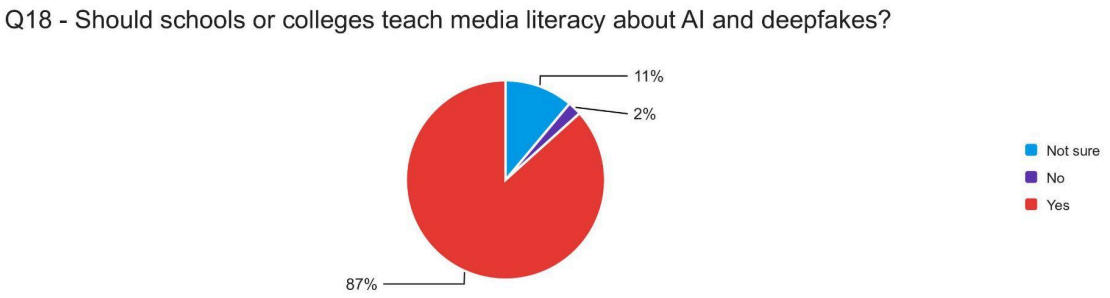


Figure 7 illustrates participants' opinions regarding whether or not schools should promote digital literacy, especially about AI and deepfakes as we transition to a time where AI becomes more prominent.

Figure 8

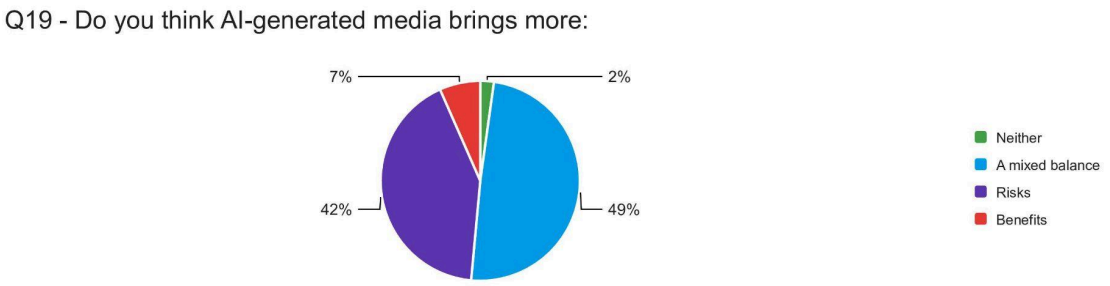
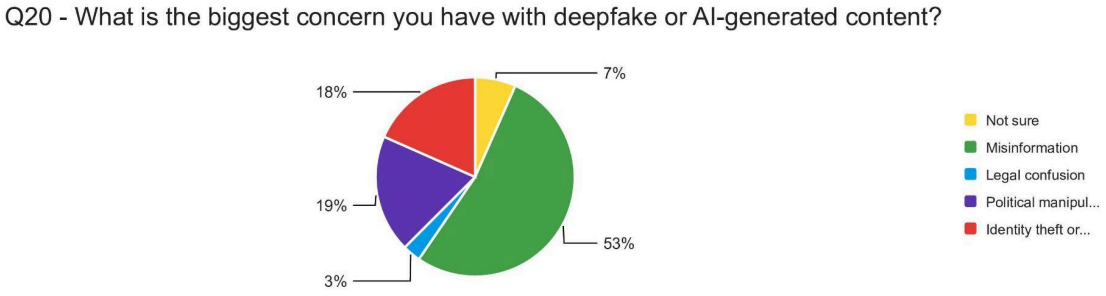


Figure 8 puts into perspective the attitudes that people have towards the implementation of AI. Roughly half of participants answered that AI content brings a mixed balance to society.

Figure 9



The survey question in *Figure 9* exemplifies public opinions of concerns with AI-generated media. Over 50% of participants worry about widespread misinformation resulting in a plethora of negative implications.

Figure 10

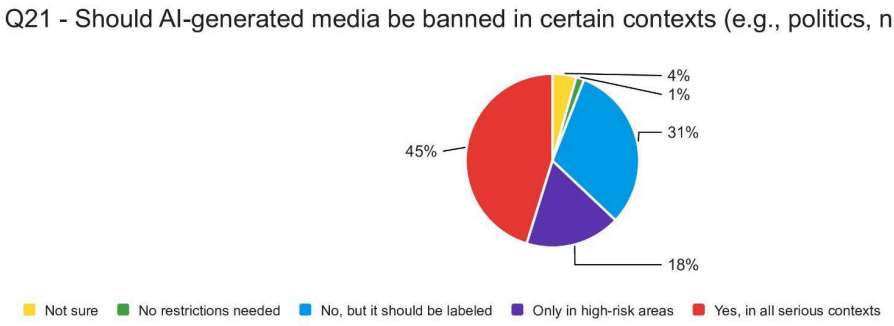


Figure 10 indicates that over 60% of individuals push for a ban of AI content in certain contexts whereas about 30% answered in favor of no necessary bans. It is important to note that a clear majority of the 30% of those in favor of no ban answered that AI content should be labeled as such.

Discussion

The results of this survey are staggering. Based on the inputs of over 150 total participants, their ability to recognize AI-generated content is strongly correlated with the quality of the generation tool. *Figure 1* illustrates how more than 75% of respondents were at least slightly confident in their ability to detect fabricated media. According to *Figure 2*, this confidence may stem from familiarity with Artificial Intelligence tools, as over 80% of participants have used such instruments to generate non-human content. *Figures 3 & 4* show that more than half of respondents feel a deep concern with the implementation of AI in legal and sociopolitical contexts. AI tools have already been used to imitate former and current Presidents of the United States (Bond 2024). Despite a range of varying concerns, it appears that an overwhelming majority of over 90% of respondents believed that AI-generated media should carry disclaimers or should, at the very least, be labeled as AI content, as per *Figure 5*. Furthermore, *Figure 7* highlights how participants feel about including digital literacy in school curriculum. Being in the 21st century and with the current expansion of technology, it has become more obvious that digital literacy should be a basic skill, similar to basic arithmetic and traditional literacy. 87% of participants feel that digital literacy should be an integral part of the educational system to ensure that today's students can be function citizens in the future. *Figure 8* covers respondents' opinions after viewing the short clips about the risks and benefits of AI. 49% believed that AI brings a mixed balance of risks and benefits to society while 42% answered that it brings more risks than benefits against the 7% that argued the contrary. This is imperative because, after viewing the video recordings, about six times as many people responded that AI brings more risks than people who responded that AI brings more benefits. *Figure 9* indicates that the major concern that individuals have regarding AI content is the spread of

misinformation. The development and improvements to hyper-realistic deepfakes opens more doors to the manipulation and spread of misinformation (Given 2025). *Figure 10* explains how most of the individuals that took part in the survey opted for some sort of ban on AI in certain contexts like social media, politics, and the news. Only about 32% stated that AI should be unregulated or should at least be labeled as AI-generated content for safety and privacy reasons.

It is also crucial to understand that videos 1 & 3 were both AI-generated, but at different qualities. Most people were certain that the first video was AI-generated, yet for the third video clip, it was surprising that most participants were certain that it was real content. This shows that a certain quality of AI audio and video can bypass human perceptual skills. As time passes, AI will continue to aggregate digital data about humans, updating and improving its code, and produce better and better simulations of human speech and behavior. If, at this point, people shakily and inconsistently identify deepfakes, a future with unregulated control over the capability of AI will ensue, influencing social, political, and even economic contexts. An initiative must be taken to address the lack of digital literacy and the poorly regulated privacy concerns and protections of individuals when it comes to the introduction of AI. Not only do AI deepfakes threaten to destabilize socio-political and economic contexts, but also threaten people on an individual level with defamation attacks, crushing the reputations and destroying livelihoods.

Limitations

After conducting the survey, there are some changes I would have made to the process. Primarily, I would have preferred to publish the survey earlier to ensure that as many people can fill it out as possible. I also would have shared the survey in public spaces so that the people taking the survey are not only friends and family, but also strangers throughout campus and their extended friends and family, broadening the range of data that can be collected. Moreover, I would have included several other questions, excluded some questions, and perhaps added one or two more video clip examples to get more details about the content itself. Finally, I would have also liked to expand and specify the multiple choice answer choices in each survey question to eliminate any confusion and uncertainty that may have come about.

Conclusion

This critical analysis highlights a growing gap between the realism of synthetic media and the public's ability to recognize it. As deepfakes become growingly advanced and readily convenient, the risks to privacy, public trust, and digital safety increase. The results from the survey emphasize the need for digital literacy, platform accountability, and clear labeling around AI content. Left unchecked, the line between what's real and what's fake will continue to blur, leaving people vulnerable to manipulation and cyber attacks. Future research should incorporate a larger sample population and ways to assess AI-generated content for inconsistencies to improve the ability to recognize AI content in this changing digital world.

References

"Ai Deepfake Security Concerns." CSA. Accessed April 15, 2025. https://cloudsecurityalliance.org/blog/2024/06/25/ai-deepfake-security-concerns.

Ainsworth, Amber. "'A Social Problem': Growing Trend of AI Deepfakes and Misinformation Spurs Fears as Technology Advances." FOX 2 Detroit, October 28, 2024. https://www.fox2detroit.com/news/a-social-problem-growing-trend-ai-deepfakes-misinformation-spurs-fears-technology-advances.

Bond, Shannon. "How AI Deepfakes Polluted Elections in 2024." NPR, December 21, 2024. https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections.

CBS News. 2024. "Could You Spot a Deepfake Video? A Boston Area Survey Showed More than Half Failed the Test." CBS News. Accessed April 15, 2025. https://www.cbsnews.com/boston/news/i-team-deepfake-videos-artificial-intelligence/.

Deceptive audio or Visual Media ("deepfakes") 2024 legislation. Accessed April 16, 2025. https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation.

"Deepfake Technology." Organization for Social Media Safety, December 20, 2019. https://www.socialmediasafety.org/advocacy/deepfake-technology/.

"Deepfakes and the Ethics of Generative AI." Tepperspectives, August 19, 2024. https://tepperspectives.cmu.edu/all-articles/deepfakes-and-the-ethics-of-generative-ai/.

Given, Lisa M. Professor of Information Sciences & Director "Generative AI and Deepfakes Are Fuelling Health Misinformation. Here's What to Look out for so You Don't Get Scammed." The Conversation, March 13, 2025. https://theconversation.com/generative-ai-and-deepfakes-are-fuelling-health-misinformation-heres-what-to-look-out-for-so-you-dont-get-scammed-246149.

National Center for State Courts (NCSC). 2024. "Digital Evidence and Deepfakes in the Age of Ai." National Center for State Courts. Accessed April 16, 2025. https://www.ncsc.org/__data/assets/pdf_file/0019/101683/ncsc-ai-rrt-deepfakes-june-2024.pdf.

"How to Spot Deepfake Scams." What Are Deepfake Scams and How Do I Spot Them? Accessed April 15, 2025. https://www.ncoa.org/article/understanding-deepfakes-what-older-adults-need-to-know/.

"Language Selection." Fairfax County Virginia. Accessed April 15, 2025. https://www.fairfaxcounty.gov/familyservices/older-adults/golden-gazette/2024-03-artificial-intelligence-and-deepfake-videos-what-you-need-to-know.

Saul, Derek. 2024. "Deepfake Video Convinces Firm to Pay $25 Million." *Forbes*. https://www.forbes.com/sites/dereksaul/2024/02/07/deepfake-video-convinces-firm-to-pay-25-million-in-massive-scam/

Text - H.R.5586 - 118th Congress (2023-2024): Deepfakes accountability act | congress.gov | library of Congress. Accessed April 16, 2025. https://www.congress.gov/bill/118th-congress/house-bill/5586/text.