The Effect of Media Literacy on Misinformation and Deep Fake Video Detection

Sarah Shawky El Mokadem[•]

Abstract

These days, misinformation is a major problem in virtually all societies. The emergence of deepfake AI technology has only exacerbated issues around misinformation and disinformation. One way to try to limit the negative effects of fake news and false information in a society is media literacy. This study focused on the effect of media literacy on detecting fake news in both video and text messages. Based on Inoculation Theory, three hypotheses were suggested and a quasi-experimental research design was conducted. The participants included 204 highly educated Egyptians who were divided into three groups. One group was given a powerpoint lecture on the meaning of misinformation, its consequences, how to verify information received via social media, and examples of fake news. The second group was given a similar presentation but instead of talking about misinformation in general the presentation focused on deepfake AI technology. The third group was the control and were not given any media literacy information. Two types of fake messages were evaluated by each group; a Facebook post with text and an image and a deepfake video. The results of the study suggest the effectiveness of media literacy lectures in raising the rate of detection of fake news whether in the form of text or video. Results also indicated a decreased probability of sharing the fake messages. The study recommends designing media literacy lectures including both information on misinformation in general and specific information on the new technology creating fake news like deepfake.

Introduction

The main function of the media is to inform, but since the emergence of the first news outlet the problem of misinformation has existed. Rumor, frequently linked with hearsay, gossip, or informal communication, could be considered the most ancient member of the misinformation lineage (Adams

---

[•] Assistant Professor in the Department of Mass Communication, Faculty of Al Alsun and Mass Communication.
Misr International University

et al. 2023). With the spread of social media, information circulates much faster and reaches a larger audience. Although social media has great potential for spreading awareness and decreasing knowledge gaps, it has made easier the creation and spread of misinformation (Hilary and Dumebi 2021), which consequently transformed the misinformation problem into a major crisis warranting brisk action (Mason, Krutka and Stoddard 2018).

The problem of misinformation in the media is not new, however was brought to the fore with the emergence of the coronavirus pandemic; creating a worldwide vaccine scare (Sharma, Zhang and Liu 2022; Ahmadi 2022; Skafle, et al. 2022). Misinformation entails conveying inaccurate or untrue details. Instances of this encompass unfounded stories, offensive language, and practical jokes. Disinformation, however, is a purposeful act that involves sharing harmful content like fraudulent schemes, targeted deceptive emails, and manipulative propaganda. This deliberate dissemination generates feelings of fear and doubt within the populace (Using Social Media in Community Based Protection: A Guide 2022). Another relatively new term related to the problem of misinformation is fake news. The term fake news refers to information presented in a news-like manner that is deliberately and demonstrably untrue. This distinction sets it apart from satire, parody, and propaganda (Adams et al. 2023). The main difference between these terms is the intent. Nevertheless, the consequence of spreading misinformation is as dangerous as the consequence of spreading disinformation.

The main problems of misinformation, disinformation, fake news are threefold; it goes viral quickly, it creates the phenomenon of "continued influence effect", and the phenomenon of "illusory truth effect" (Roozenbeek and Linden 2021). First, misinformation, disinformation, and fake news spread faster than fact-checked information, since fact checking takes time. Second, even after the misinformation, disinformation, or fake news is rebuked, audiences often still defer to them if they are the first information that was consumed, and sometimes the factual information never reaches them. Lastly, if misinformation, disinformation, or fake news is spread widely enough, it is often repeated over and over, which fortifies the belief in it even if audiences know it is false (Roozenbeek and Linden 2021). This means that even if the audience knows the content they were exposed to was fake, it will still have a persuasive effect on them that will be used to interpret new information in the future. This is called the "sleeper effect" (Kietzmann, Mills, and Plangger 2021).

The spread of untrue, false, or inaccurate information deliberately is not limited to people wanting to affect or control public opinion. Sometimes social media users are afraid to be left out if they do not share the fake information they receive (fear of missing out of "FOMO") (Ahmed 2022).

**The Effect of Media Literacy on
Misinformation and Deep Fake Video Detection**

Some audiences perceive sharing fake information and videos as a harmless act that everyone is doing.

Despite the evidence of the spread of misinformation, disinformation, and fake news, it is not easy to convince users to be cautious about information they receive through social channels. One reason is that users do not believe that they could be deceived by fake news (Jang and Km 2018). Furthermore, today's hectic news scene does not allow for sufficient screening of the news before spreading it. This has led to the beginning of a new era called the "post-truth" era (Westerlund 2019), where people tend to believe the "realities" they want to believe rather than objective truth (Yildirim and Yolcu 2022).

The spread of these phenomena have also created another problem, which is excessive mistrust among media consumers. Increasingly, audiences are hesitant to believe any type of information, even from credible sources, often assuming there is a "conspiracy". This includes questioning the reality of serious global issues such as global warming (Hameleers 2022; Dan et al. 2021).

The role of media literacy is to inform audiences on how to judge the credibility of information they receive on a daily basis from all sources and to create a balance between trusting and being skeptical of the news. One way of doing this is helping audiences detect false information.

Literature Review

This study aims at examining how media literacy can mitigate the effect of misinformation and deepfake technology in the digital era. The review explores the credibility disparity between visual and textual messages, delving into the effects of deepfake technology and the critical role of media literacy in mitigating its impact on misinformation, disinformation, and fake news. It also investigates how Inoculation Theory can be applied to enhance media literacy and protect individuals against falling prey to fake media messages, thus safeguarding the trustworthiness of information.

Credibility of visual vs. textual messages and the effect of deepfakes

Visual messages are more persuasive than textual messages because they offer an additional layer of proof. Additionally, visuals are easier to understand by a larger audience (Dan et al. 2021). Visual messages also elicit emotional reactions, which make them more effective in changing people's attitudes and behaviors (Dan et al. 2021; El Mokadem 2018) while also affecting how users choose and perceive information (Sikorski et al. 2021). For this reason, deepfake videos pose a significant threat in spreading false information quickly.

The term "Deepfake", which means the use of artificial intelligence or AI to create "hyper-realistic" videos of people saying and doing things that never

happened, was coined in 2017 (Westerlund 2019). This manipulation of reality affects many aspects of life, but until now the use of this technology has mainly focused on entertainment, politics, and arts (Yildirim and Yolcu 2022; Sikorski et al. 2021).

Deepfake technology has added new dimension to the problem of impersonation on social media (Popescu 2020). It is no longer restricted to creating fake accounts or using other people's profile pictures, but rather creating videos with people saying things that never happened. Despite audiences believing they could detect deepfake videos when exposed to them, the technology is very hard to detect (Thaw, et al. 2021). Deepfake technology has shattered the idea of "seeing is believing" (Shin and Lee 2022). Videos are regarded as more credible than pictures, which in turn are more credible than textual messages. Furthermore, a video can easily trigger an emotional response as it is more believable than text or sound. Deepfake videos are more vivid and trustworthy than fake textual messages or images, which makes them more dangerous (Hwang, Youn, and Jeong 2021). Until the general public is able to discern real information from false, media literacy is an important factor in decreasing the spread of misinformation, disinformation, and fake news, including deepfake videos.

However, it is not all dark and gloomy in predicting the future of news with the development of deepfake technology. There are other factors that may affect the perception of deepfake video (and other forms of false information) and limiting its spread. One experiment conducted in the Netherlands concluded that deepfake messages did not have a more credible stance among the target audience. In this study, a more powerful factor to predict the detection of misinformation is the personal biases of the target audience. People who already agreed with the misinformation evaluated these messages as credible (Hameleers, Van Der Meer, and Dobber 2022).

This could be explained by the results of another study that found that when the target audience became more knowledgeable about a candidate's personality, they were able to spot deepfake videos more often (Ng 2022). This supports the hypothesis that the more knowledgeable the audience, the more effective they are at discerning fake media messages.

In recent years, due in large part to the prevalence of false information on social media, mistrust in any media news and information as a whole is growing. A study by Michael Hameleers (2022) suggested that researchers should not assume respondents have a "univocal acceptance" of neutral information. The results of this research concluded that users who are already distrusting the news found the news from even well-established news organizations untrustworthy (Hameleers 2022).

Misinformation and deepfake problems can arguably be diminished through legislation and education (Westerlund 2019; Nurcahyani, Junaedi, and

**The Effect of Media Literacy on
Misinformation and Deep Fake Video Detection**

Sudiwijaya 2022). A number of researchers measured the role of media literacy on diminishing the tendency to share videos, thus limiting deepfake videos from going viral. One such study reached the result that recognizing deepfake videos leads to decreased sharing and therefore limiting its negative effect (Iacobucci et al. 2021).

Media literacy and Inoculation Theory

Inoculation Theory is a social psychology theory on persuasion. The main idea behind this theory is to make attitudes resistant to change (Compton 2012). The theory was developed in the 1960s by McGuire and was mainly applied to health attitudes (Compton, Jackson, and Dimmock 2016). Later on it was used in politics, public relations, and even media (Breen and Matusitz 2009). The idea was that, like a vaccine, messages can be used to make attitudes resistant to change. For example, in an anti-smoking campaign, the messages could be directed to youth who aren't smokers and haven't been exposed to any persuasion to make them start smoking. A weaker version of the arguments that can be used to convince them of smoking are introduced with the counter arguments of how to respond to them. So, when the youth are faced with these arguments in real life, they are already "vaccinated" against them and thus more resistant to persuasive attempts to make them smoke (Compton 2012; Severin and Tankard 2014).

There are two ways to make attitudes resistant to change, the supportive method and the inoculation method. The supportive method depends on focusing on the existing positives and making audiences cling to these aspects (Severin and Tankard 2014).

The process of achieving successful inoculation involves three stages: (1) the warning, (2) the weak attack, and (3) the active defending (Breen and Matusitz 2009). The warning stage of the inoculation process involves notifying participants about an upcoming argument and advising them to be prepared for a counter attitudinal attack. In the subsequent stage, referred to as the weak attack, participants are presented with a mildly challenging argument designed to be easily rejected. Finally, in the active defending stage, participants are tasked with effectively defending their beliefs using simple defense mechanisms. Together, these stages synergistically contribute to the overall effectiveness of the inoculation process (Breen and Matusitz 2009).

There are two strategies to inoculate attitudes and make them resistant to change, refutational-same and refutational-different (Compton et al. 2021). Refutational-same strategy offers counterarguments specific to future attacks, while refutational-different strategy offers counterarguments not specific to future attacks, but similar (Compton, Jackson, and Dimmock 2016).

The same concept could be applied on media literacy. Media literacy is a concept that appeared in the 1980s after the extensive spread of media (Neag, Bozdağ, and Leurs 2022). Media literacy teaches the audience to be critical and to question all messages they receive through any form of media. It encourages audiences to check sources of information and to verify the information itself from more than one source. Thus, media literacy is used to make audiences resistant to believing fake news.

In recent years, there is an interest in how Inoculation Theory could be used to reduce misinformation, disinformation, and fake news on one hand and reduce the effect of conspiracy theories on the other (Compton, Linden, et al. 2021; Roozenbeek and Linden 2021). Media literacy can be used to "inoculate" users and increase their ability to identify false information (Jones-Jang, Mortensen, and Liu 2019). It is used to make people more resilient to believing misinformation, disinformation, and fake news, rather than using "reactive resources" (McDougall 2019).

Based on Inoculation Theory, this study proposes the following hypothesis:

H1: The respondents who did not receive any lecture will rate the video message and the text message as credible more often than those who received a media literacy lecture.

The problem with inoculation is that the effect does not last long (Severin and Tankard 2014). It may only last as little as two weeks (Banas and Rains 2010). One idea to overcome this problem is offering a booster message to resist the fading effect of the inoculation messages, but the effect of these booster messages is mild at best (Ivanov, Parker, and Dillingham 2018; Compton 2012). Another problem is that sometimes, like a vaccine, the counter arguments effect does not work with new or novel arguments (Severin and Tankard 2014). That is why some research has focused on the effect of refutational-different strategy; sometimes seen as a more effective way to combat novel attacks. However, the results are still not conclusive (Banas and Rains 2010). Based on that, this study proposes the following research questions:

> *RQ1: What are the effects of a deepfake media literacy lecture vs. a misinformation lecture on the evaluation of credibility of both types of messages?*

Variables and Definitions:

The independent variable for this study is the type of media literacy lecture. It is a categorical variable including misinformation lecture, deepfake lecture, or no lecture.

- The independent variable is the media literacy lecture.

- The dependent variable is the detection of misinformation or deepfake videos.

- The intermediate variable is the type of the message whether it is written or video.

The choice of the term misinformation over disinformation or fake news is driven by the researcher's intention. Given that the aim is not to deliberately cause harm and the participants were made aware that the news they received was false, misinformation better suits the purpose of this research.

Definitions of variables:

1- *Detection of misinformation/deepfake videos:*

- Conceptual definition: According to Oxford Languages, detection (noun) is the action or process of identifying the presence of something concealed (Oxford Languages n.d.).

- Operational definition: In this study the detection of misinformation was measured using the rating of participants to the credibility of the message.

2- *Credibility of the message:*

- Conceptual definition: According to Oxford Languages, credibility (noun) is the quality of being trusted and believed in, or the quality of being convincing or believable (Oxford Languages n.d.), according to the Britannica Dictionary, it means the quality of being believed or accepted as true, real, or honest (The Britannica Dictionary n.d.), and according to Cambridge Dictionary, credibility (adjective) means able to be believed or trusted (Cambridge Dictionary n.d.).

- Operational definition: In this study, the researcher used a 7-point scale where participants rated the message they received based on its perceived credibility, persuasiveness, accuracy, truthfulness, and whether the message is convincing to them.

Methodology

This research is quasi-experimental. Quasi-experimental studies have higher internal validity than correlational research, but a lower internal validity than experimental studies. This methodology was chosen to assign participants to specific groups according to their age groups, knowledge of deepfake technology and their perception of social media credibility. In addition to that, for practical reasons, the quasi-experimental research design was chosen to study participants in a semi-natural environment (as a part of their curriculum or in an informative meeting). This somewhat real setting makes the external validity of quasi-experimental research design higher than experimental research design.

The type of quasi-experimental methodology used was a combination between nonequivalent groups design and a pretest-posttest design. Quasi-

experimental research designs do not have to include a control group, but in this study a control group was included to compare between those who were given a media literacy lecture and those who were not. The study followed a 2 (deepfake video message vs. textual misinformation message) by 3 (misinformation media literacy lecture vs. deepfake media literacy lecture specifically vs. no literacy) design. By choosing this design, the experiment compares two types of media literacy strategies, which aligns to the two strategies of Inoculation Theory; refutational and supportive. The absence of media literacy in the third group serves as the experimental control. Additionally, this design facilitates the examination of textual and visual messages, as well as the respective impacts of media literacy on each.

Population and sample:

The population of this research included students and staff from a private university and was divided into two groups. The first group consisted of senior mass communication students (ages from 20 to 25). The researcher chose mass communication students as they already study the media and should be knowledgeable in detecting misinformation in media messages. The sample was purposive as the study was limited to senior students who have completed four years studying mass communication and thus considered the most knowledgeable and highly educated group among their peers.

The second group included staff (all aged above 40 years) working in the faculties of Mass Communication, Al Alsun (Language), Pharmacy, Dentistry, and Business. The population was not limited to mass communication staff alone due to the small number of staff members. The researcher tried to overcome the limitation created by the diverse specializations by purposely choosing highly educated staff with PhD degrees. The computer science staff were excluded due to their expert knowledge of deepfake technology.

The researcher divided the sample into three main groups. Half of each group consisted of students and the other half of university staff. This division was done to control the effect of the educational level and age on the experiment's results.

The majority of participants were female. In this university, the mass communication department and the university staff members primarily consist of females.

**The Effect of Media Literacy on
Misinformation and Deep Fake Video Detection**

Table (1): The division of experiment participants

Main categories	Misinformation lecture		Deepfake lecture		No media literacy lecture	
No. of participants	68		68		68	
Sub categories	Post	video	Post	video	Post	video
No. of participants	34	34	34	34	34	34
Level two subcategories	Stu.	Staff	Stu.	Staff	Stu.	Staff	Stu.	Staff	Stu.	Staff	Stu.	Staff
No. of participants	17	17	17	17	17	17	17	17	17	17	17	17

Procedures

The research was conducted on university students and staff during the month of December 2022. Three senior Mass Communication students' classes, where the number of students were 40, 37, and 35 (total of 112 students), and three university staff meetings, where the number of staff attending was 50, 38, and 35 (total of 123 staff members), were chosen to conduct the experiment. Both students and staff were informed that they were participating in an academic experiment on social media and verbal consent to participate was given to the researcher before starting. The participants were not given the specifics of the topic until after the procedures.

Before starting, a short survey was distributed to gather basic information on the respondents. They were asked about the time spent using social media daily, their favorite social media platform, their rating for social media credibility, their news sharing habits, and their knowledge of deepfake technology.

During one class and one meeting a misinformation literacy lecture was given. During the second class and meeting a deepfake literacy lecture was given. As for the third class and third meeting, no lectures were given.

After the lectures, half of the class or meeting members were given a fake piece of information in the form of a Facebook post with one photo created by the researcher (see Appendix A), and the other half were made to watch a 16 second deepfake video. Both pieces of news were about Mark Zuckerberg, the post featured a photo of Mark with a headline on the losses he suffered after a video of him claiming to use Facebook users' personal data was released. The 16 second video was an edited version of a deepfake video of Mark saying that he did use Facebook users' personal data. The video was obtained from YouTube and trimmed by the researcher (see Appendix B). The participants who were asked to evaluate the written post were given 16 seconds to look at it; the duration of the video.

The misinformation and deepfake media literacy lectures were designed based on the foundation of Inoculation Theory. First, the terms were

defined. In the misinformation lecture, the terms misinformation and disinformation were defined; and in the deepfake lecture the term deepfake was defined and its origins explained. Second, an element of threat was introduced by explaining the consequences of spreading misinformation/spreading deepfake videos. This element of threat was added as an "explicit forewarning" (Compton et al. 2021) to increase the effectiveness of the inoculation media literacy lectures (Breen and Matusitz 2009). The third part of the lectures was focused on examples of misinformation posts/deepfake videos and how to detect misinformation/deepfake videos and how to corroborate the information the users receive through social media. This part was designed based on the refutational strategy of Inoculation Theory. Most of the research found that the refutational strategy is more successful in making attitudes resistant to change in the future (Compton et al. 2021). Participants were then asked to evaluate the message they received according to their credibility, and then to state whether they would share this piece of information or not and the reasons for their decision. After the experiment, all participants were informed that the post and the video were fake messages.

Measurements

The messages were evaluated based on the measurements suggested by Hwang et al. (2021) where study participants were asked to evaluate the messages they watched on a scale from 1 to 7. Vividness was measured by three categories; whether the message was easy to picture, detailed, and specific. Credibility was measured by whether the message was accurate, credible, trustworthy, and truthful. Finally, persuasiveness was measured by whether the message was persuasive and convincing. In this study, the researcher used only the credibility category and added the convincing sub category in lieu of trustworthiness. This was done to avoid confusion in meaning among participants after conducting a pre-test on five individuals to determine the clarity of measurements.

Results

The researcher utilized SPSS for all statistical analyses, employing basic tests like frequencies and crosstabs. To examine the differences among the six groups in the study, a one-way ANOVA was employed. These differences included: (a) participants who received the misinformation media literacy lecture and evaluated the written message, (b) participants who received the misinformation media literacy lecture and evaluated the video message, (c) participants who received the deepfake media literacy lecture and evaluated the written message, (d) participants who received the deepfake media literacy lecture and evaluated the video message, (e) participants who didn't receive any media literacy lectures and evaluated the written message, and

finally (f) participants who didn't receive any media literacy lectures and evaluated the video message.

According to the Qualtrics XM website, ANOVA, also known as Analysis of Variance, was originally developed by Ronald Fisher in 1918 and has been widely used ever since. Its primary purpose is to determine whether there are significant differences between the means of three or more independent groups. The simplest form of ANOVA is referred to as one-way ANOVA (What is ANOVA? n.d.). Furthermore, after conducting the ANOVA test, the researcher employed the Tukey post-hoc test to interpret the relationships between variables. The choice of the Tukey post hoc test was based on assessing the homogeneity of variances, which did not reveal any differences between variables, indicating their homogeneity.

Sample characteristics

The total number of study participants was 235. The pre-experiment survey had a question about participants' knowledge of deepfake videos. Those who have seen deepfake videos before were excluded from the experiment. The experiment was done on the rest of the participants, a total number of 220. During analysis, and in order to have an equal number of participants in each of the groups under study for the comparison, some participants were excluded to even the number of each sub-group to 34 participants, 17 from the students and 17 from the staff members. The excluded participants were chosen based on their evaluation of social media credibility in the pre-experiment survey. The highest and lowest evaluations were excluded. The final total number of participants was 204.

Out of the 204 only 10 were males and 194 were females. The participants were divided into two age categories (aged between 20-25 and aged above 40) both with 102 participants. The ages ranged between 21 and 57 years old (mean=35.7, standard deviation=1.22, and range 42 years).

More than two thirds of the participants (71.1%) are moderate users of social media (between 2 and 5 hours a day), while 19.6% are light users (less than 2 hours daily) and only 9.3% are heavy users (more than 5 hours daily).

Table (2): Characteristics of participants in numbers

Participants	Gender		Age		Mean of social media credibility	Time spent on social media			Heard of deepfake technology	
Media literacy	Male	Female	20-25	Above 40		Light	Moderate	High	Yes	No
Deepfake	2	66	34	34	5.5	8	46	14	10	58
Misinformation	2	66	34	34	5.8	14	52	2	27	41
None	6	62	34	34	5.4	18	47	3	14	54
Total	10	194	102	102		40	145	19	51	153

The most used social media platform by participants was Facebook with 41.2% of participants stating that they use it the most, followed by Instagram with 35.3%, followed by WhatsApp with 10.8%, and finally

TikTok and Twitter with 3.4% for each. The main source of information or news used by the participants was social media (78.4%), followed by television (8.3%), then online or printed newspapers (7.8%) and finally other internet sources, like Google (5.4%). The main reasons for sharing posts or videos on social media for the participants were when they feel the post or video expresses their emotions, experiences, or inner thoughts (26%) or when they believe it is an important or useful piece of information (21.1%) or when they see it as funny (20.1%).

Only one quarter of the participants had heard about deepfake videos (26 participants over 40 years of age and 25 participants aged between 20-25).

H1: The respondents who did not receive any lecture will rate the video message and the text message as credible more often than those who received a media literacy lecture.

The ANOVA test results show that there is a significant difference between the six groups of the study in terms of evaluating the persuasiveness, accuracy, and whether the message is convincing at 0.001 significance level. There is also a significant difference between the six groups of the experiment in terms of the evaluation of the message credibility and truthfulness at significance level 0.005 and 0.05 respectively.

Table (3): ANOVA results comparing the means of the six groups under study

		Sum of Squares	df	Mean Square	F	Sig.
the message was persuasive	Between Groups	71.451	5	14.290	5.256	.000
	Within Groups	538.294	198	2.719		
	Total	609.745	203			
the message was convincing	Between Groups	87.157	5	17.431	5.998	.000
	Within Groups	575.471	198	2.906		
	Total	662.627	203			
the message was credible	Between Groups	58.451	5	11.690	3.467	.005
	Within Groups	667.588	198	3.372		
	Total	726.039	203			
the message was truthful	Between Groups	38.672	5	7.734	2.551	.029
	Within Groups	600.324	198	3.032		
	Total	638.995	203			
the message was accurate	Between Groups	56.162	5	11.232	4.366	.001
	Within Groups	509.382	198	2.573		
	Total	565.544	203			

The Tukey post-hoc test results indicate that the control group, those who did not receive any media literacy lectures, rated both types of messages higher in terms of persuasiveness, credibility, truthfulness, and accuracy compared to participants who received any type of media literacy lectures. This generally supports the main research hypothesis.

Further examination of the results shows variations in the evaluation of the

five credibility categories based on the type of message evaluated. Specifically, the control group's results exhibit differences between the evaluation of written and video messages. Participants' assessment of the persuasiveness and accuracy of the written message, as well as how convincing it is, was the highest among all six groups, with means of 4.3824, 3.9412, and 4.5882, respectively. Conversely, the video message received the highest ratings from participants who did not receive any media literacy lectures, particularly in the credibility and truthfulness categories, with means of 4.1471 and 3.7941, respectively.

These findings raise questions about the influence of message type, whether written or visual, on its credibility evaluation, underscoring the overall importance of media literacy in detecting misinformation and deepfake videos. In conclusion, the results support the validity of hypothesis H1.

RQ1: What are the effects of a deepfake media literacy lecture vs. a misinformation lecture on the evaluation of credibility of both types of messages?

The study's results indicate that participants who received the deepfake lecture rated the credibility of the video message the lowest among all six groups in four of the five evaluation categories. The mean scores for credibility, accuracy, truthfulness, and being convincing were 2.6176, 2.4706, 2.6176, and 2.7353, respectively. However, interestingly, the only category where the evaluation of the video message among participants who received the deepfake lecture was not the lowest was the persuasiveness of the message. In this category, the participants who received the misinformation media literacy lecture and evaluated the video message rated it the lowest, with a mean score of 2.8235, and the deepfake lecture group evaluating the video message was a close second, with a mean score of 2.8824.

On the other hand, those who received no media literacy lectures at all rated the credibility and truthfulness of the video message the highest among the six groups, with mean scores of 4.1471 and 3.7941, respectively. Conversely, the written message received the highest evaluation of accuracy, persuasiveness, and being convincing from participants who did not receive any media literacy lectures, with mean scores of 3.9412, 4.3824, and 4.5882, respectively.

Table (4): Tukey HSD Multiple Comparisons

Category of evaluation	Groups under study	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Credible	Deepfake lecture and video message vs. no lecture and video message	-1.5294	.44535	.009	.2478	2.8110
	Deepfake lecture and video message vs. no lecture and written message	-1.20588	.44535	.05	-.0758	2.4875
	Deepfake lecture and video message vs. misinformation lecture and written message	-1.20588	.44535	.05	-2.4835	.0758

Category of evaluation	Groups under study	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Persuasive	No lecture and video message vs. deepfake lecture and video message	1.32353	.39990	.014	.1727	2.4744
	No lecture and video message vs. misinformation lecture and video message	1.38235	.39990	.007	.2315	2.5332
	No lecture and written message vs. deepfake lecture and video message	1.50000	.39990	.003	.3491	2.6509
	No lecture and written message vs. misinformation lecture and video message	1.55882	.39990	.002	.4080	2.7097
Truthful	No lecture and video message vs. deepfake lecture and video message	1.17647	.42231	.05	-.0389	2.3918
	No lecture and written message vs. deepfake lecture and video message	1.17647	.42231	.05	-.0389	2.3918
Convincing	No lecture and video message vs. deepfake lecture and video message	1.44118	.41348	.008	.2512	2.6311
	No lecture and video message vs. misinformation lecture and video message	1.23529	.41348	.037	.0454	2.4252
	No lecture and written message vs. deepfake lecture and video message	1.85294	.41348	.000	.6630	3.0429
	No lecture and written message vs. deepfake lecture and written message	1.17647	.41348	.05	-.0135	2.3664
	No lecture and written message vs. misinformation lecture and video message	1.64766	.41348	.001	.4571	2.8370

Furthermore, the statistical analysis, as shown in Table 3, indicates significant differences between the participants who received the deepfake media literacy lecture and evaluated the video messages, and those who received the misinformation lecture and evaluated the written message when it comes to evaluating the credibility of the messages, with a significance level of 0.05. There is also a significant difference in evaluating the message credibility between participants who received the deepfake media literacy lecture and evaluated the video messages, and those who received no media literacy lectures, whether they evaluated the written message (at a confidence level of 0.05) or the video message (at a confidence level of 0.01). The deepfake lecture had a more substantial effect on participants evaluating the video message compared to those who received the misinformation media literacy lecture and evaluated the written message. These results suggest that written messages require more media literacy lectures to be as effective or focused as deepfake media literacy, although this might prove challenging.

For the evaluation of the persuasiveness of the messages, there were significant differences between those who received the deepfake media literacy and evaluated the video message, and those who received no lecture and evaluated both the video message and the written message, with

confidence levels of 0.05 and 0.005, respectively. Additionally, there were significant differences between those who received the misinformation lecture and evaluated the video message, and those who received no lecture and evaluated both the video message and the written message, with confidence levels of 0.01 and 0.005, respectively. These findings indicate that media literacy lectures effectively reduced the percentage of participants persuaded by both the written and video messages.

Regarding the evaluation of the truthfulness of the messages, there was a significant difference between those who received the deepfake media literacy and evaluated the video message, and those who received no lecture and evaluated both the video message and the written message, with a confidence level of 0.05. However, this difference was not observed for those who received the misinformation lecture, indicating that, for this category, the deepfake media literacy lecture was more effective than the misinformation lecture.

Finally, when evaluating whether the message was convincing to the participants or not, the deepfake media literacy lecture proved to be more effective. There were significant differences between the participants who received the deepfake media literacy and evaluated the video message, and those who received no lecture and evaluated both the video message and the written message, with confidence levels of 0.01 and 0.001, respectively. Interestingly, participants who received the misinformation media literacy lecture evaluated the video message as less convincing compared to those who evaluated the written message.

These results suggest that deepfake media literacy has a more pronounced effect on the evaluation of the credibility of the video message compared to the impact of misinformation media lecture on the video message. This indicates that if media literacy is not directly focused on deepfake videos, the video message will generally be viewed as more credible than the written message. Moreover, these findings highlight the importance of considering the type of message when evaluating its credibility, with written messages being perceived as more accurate and convincing, while visual messages, particularly videos, considered more truthful and believable.

Upon analyzing the outcomes concerning variations arising from age or educational level, no statistically significant differences were found between the impacts of media literacy presentations on students versus staff members. Thus, it can be inferred that the influence of media literacy on both misinformation and deepfake videos is consistent for both student and staff groups. This suggests that the receipt of media literacy education and the nature of the message and its content play a more telling role than age or educational level in predicting the capacity to detect misinformation and deepfake videos.

Discussion

The COVID-19 pandemic provided a glaring example of how misinformation can lead to real-world consequences, such as vaccine hesitancy. Misinformation, disinformation, and fake news spread quickly and are difficult to dispel, leading to the "continued influence effect" and the "illusory truth effect," where false information persists even after being debunked. Deepfake technology has added another layer of complexity to this problem. Visual messages are more persuasive and emotionally impactful than text, making deepfake videos a potent tool for spreading false information. Deepfake videos can be challenging to detect, further complicating efforts to combat their negative impacts.

In this context, media literacy emerges as a crucial tool to address the spread of misinformation and deepfake videos. The study's results offer support for the hypothesis proposed by Inoculation Theory, highlighting the significance and effectiveness of media literacy in combating misinformation overall. Particularly fascinating are the findings when comparing the evaluation of written and video messages, which present some contradictory aspects, leading to further questions. The study indicates that the refutational-same strategy is more effective than the refutational-different strategy when evaluating video messages, but not when assessing written messages. Notably, the deepfake media literacy lecture proved more effective in facilitating the detection of misinformation for both the written and video messages compared to the general misinformation lecture. This observation could be linked to the nature of deepfake media literacy itself. For many participants, this served as their initial exposure to the concept of deepfake videos, or perhaps even the first time they had encountered the idea. Consequently, this lack of familiarity may have prompted more cautious responses from the participants, in contrast to the well-known concept of the misinformation lecture, which was more familiar to most individuals in the study.

Additionally, it might be related to the inherent credibility of video messages, which tend to be more readily accepted by audiences (Dan et al. 2021; Sikorski et al. 2021; El Mokadem 2018). Therefore, participants who received either the general misinformation lecture or no media literacy lecture at all were more likely to perceive the video message as more credible and truthful than the written message. The ongoing debate regarding the effectiveness of the refutational-same versus refutational-different strategy (Compton et al. 2021; Compton, Jackson, and Dimmock 2016) further underscores the importance of these results in enhancing our understanding of both strategies' efficacy in combating misinformation.

While some findings align with previous research (Shin and Lee 2022; Hwang, Youn, and Jeong 2021) that suggests video messages are generally

perceived as more credible and truthful due to their visual nature, other results indicate that audiences view written messages as more accurate, convincing, and persuasive. These discrepancies could be attributed to the personal biases and pre-existing beliefs of the audience concerning written versus visual messages, rather than being solely influenced by the message's actual content. This finding is consistent with the conclusions drawn by Hameleers, Van Der Meer, and Dobber (2022), who suggest that audience biases, personal experiences, and backgrounds play a significant role in how information is judged and misinformation is detected. This could also explain the reason why age and educational level had no significant effect on the experiment's results.

Ultimately, media literacy emerges as a crucial tool in empowering individuals to discern truth from falsehood and resist the spread of false information. Further research and targeted strategies may be required to ensure effectiveness for different types of messages, as in visual vs. text. Additionally, personal biases can influence how individuals evaluate the credibility of information, highlighting the need for continuous efforts in media literacy education.

Conclusion

This study aimed to explore the impact of media literacy as an inoculation strategy against deepfake videos and misinformation and investigates whether providing media literacy lectures can effectively increase resistance to deceptive messages. By shedding light on the credibility of visual versus textual messages and the role of media literacy in countering deepfakes and misinformation, this research aims to contribute to the development of strategies to mitigate the negative effects of deepfake technology on society. The study is a quasi-experimental research with higher internal validity than correlational studies but lower than experimental studies. It used a combination of nonequivalent groups and a pretest-posttest design, including a control group, to compare the effects of different media literacy lectures on participants' perception of deepfake video messages and text misinformation messages. The quasi-experimental design was chosen to assign participants to specific groups based on age, knowledge of deepfake technology, and perception of social media credibility, and to study them in a semi-natural environment for higher external validity. Results support the effectiveness of media literacy in combating misinformation, especially in relation to deepfake videos versus written messages. The refutational-same strategy proves more effective for evaluating video messages, and the deepfake media literacy lecture enhances participants' ability to detect misinformation in both written and video messages, possibly due to the novel nature of deepfake technology and the inherent credibility of visual messages. However, discrepancies in audience perceptions of credibility between written and video messages are influenced by personal biases and

pre-existing beliefs, emphasizing the role of individual backgrounds in information judgment and misinformation detection.

Limitations

1. The study's sample size is relatively small, which might limit the generalizability of the findings. Future research should aim to include a larger and more diverse participant pool to enhance the study's external validity.

2. The participants in this study were from a specific age group and geographical location, which might limit the generalizability of the results to other populations. Future research should consider including participants from various age groups and cultural backgrounds to obtain more comprehensive insights.

3. The study was conducted in a controlled experimental setting, which may not fully reflect the complexity and dynamics of real-world media consumption and information evaluation. Future research could explore the credibility evaluation of misinformation and deepfake in more natural settings.

4. The study did not consider the social media platform itself, like Facebook, Instagram, WhatsApp, Twitter, and TikTok. However, different platforms may have varying effects on the spread and perception of misinformation and deepfake content. Future research should consider investigating a broader range of social media platforms.

5. The study's duration was limited, and the participants were evaluated shortly after receiving the media literacy lectures. Future research should examine the long-term effects of media literacy interventions on the evaluation of misinformation and deepfake content.

Recommendations for Future Research

1. Conduct longitudinal studies would provide valuable insights into the long-term impact of media literacy interventions on individuals' ability to discern and resist misinformation and deepfake content.

2. Investigating the impact of cultural factors on individuals' susceptibility to misinformation and deepfake content would add further nuance to future studies, while cross-cultural studies could shed light on how media literacy approaches need to be tailored to different cultural contexts.

3. Consider Multimodal Misinformation by studying the combined impact of text, images, and videos in spreading misinformation and deepfake content. Understanding how different media types interact in influencing perceptions can lead to more effective media literacy strategies.

4. Explore how factors like emotion, cognitive load, and prior knowledge influence individuals' responses to misinformation and deepfake content.

**The Effect of Media Literacy on
Misinformation and Deep Fake Video Detection**

This could help develop targeted media literacy interventions based on specific contextual factors.

5. Conduct research on the effectiveness of media literacy interventions implemented in real-world educational and social environments. Collaborating with educational institutions and social media platforms can provide valuable insights.

6. Investigate the role that digital platforms can play in curbing the spread of misinformation and deepfake content. This includes exploring the effectiveness of AI-based content moderation and user reporting systems.

7. Examine how different age groups respond to media literacy interventions and whether there are generational differences in susceptibility to misinformation and deepfake videos.

Works Cited

2022. *Using Social Media in Community Based Protection: A Guide.* United Nations High Commissioner for Refugees.

> https://www.unhcr.org/innovation/wp-content/uploads/2022/02/Factsheet-4.pdf.

Adams, Zoë, Magda Osman, Christos Bechlivanidis, and Björn Meder. 2023. "(Why) Is Misinformation a Problem?" *Perspectives on Psychological Science* 1-28. doi:10.1177/17456916221141344.

Ahmadi, Elina. 2022. "Misinformation on Social Media Platforms in the Global Crisis of Coronavirus." In *Information Manipulation and Its Impact Across All Industries,* by Maryam Ebrahimi, 169-179. IGI Global. doi:10.4018/978-1-7998-8235-0.ch010.

Ahmed, Saifuddin. 2022. "Disinformation Sharing Thrives with Fear of Missing Out Among Low Cognitive News Users: A Cross-national Examination of Intentional Sharing of Deep Fakes." *Journal of Broadcasting & Electronic Media* 66 (1): 89-109.

> doi:10.1080/08838151.2022.2034826.

Banas, John , and Stephen Rains. 2010. "A Meta-Analysis of Research on Inoculation Theory." *Communication Monographs* 77 (3): 281-311.

Breen, Gerald Mark, and Jonathan Matusitz. 2009. "Inoculation Theory: A Theoretical and Practical Framework for Conferring Resistance to Pack Journalism Tendencies." *Global Media Journal.*

n.d. *Cambridge Dictionary.* Accessed November 2, 2022.

> https://dictionary.cambridge.org/us/dictionary/english/credible.

Campbell, Colin, Kirk Plangger, Sean Sands, and Jan Kietzmann. 2022. "Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising." *Journal of Advertising* 51 (1): 22-38.

doi:10.1080/00913367.2021.1909515.

Chudinova, Anatoly P, Natalya N Koshkarova, and Natalia B Ruzhentseva. 2019. "Linguistic Interpretation of Russian Political Agenda Through Fake, Deepfake, Post-Truth." *Journal of Siberian Federal University. Humanities & Social Sciences* 1840–1853.

Compton, Joshua. 2012. "Inoculation Theory." In *The SAGE Handbook of Persuasion: Developments in Theory and Practice*, by James Price Dillard and Lijiang Shen, 220-236. SAGE Publications, Inc.

Compton, Josh , Ben Jackson, and James A Dimmock. 2016. "Persuading Others to Avoid Persuasion: Inoculation Theory and Resistant Health Attitudes." *Front Psychol* 7 (122). doi:10.3389/fpsyg.2016.00122.

Compton, Josh, Sander van der Linden, John Cook, and Melisa Basol. 2021. "Inoculation Theory in the Post-Truth Era: Extant Findings and New Frontiers for Contested Science, Misinformation, and Conspiracy Theories." *Wiley* 1-16. doi:10.1111/spc3.12602.

Dan, Viorela, Britt Paris, Joan Donovan, Michael Hameleers, Jon Roozenbeek, Sander van der Linden, and Christian von Sikorski. 2021. "Visual Mis- and Disinformation, Social Media, and Democracy." *Journalism & Mass Communication Quarterly* 98 (3): 641-664.

El Mokadem, Sarah Shawky. 2018. "Reporting Conflict in the Middle East: A Visual and Editorial Analysis of Arabic vs. English News Channels Reports in 2017." *Scientific Journal for Radio and Television Research* 29-52.

Gamage, Dilrukshi , Jiayu Chen, Piyush Piyush Ghasiya, and Kazutoshi Sasahara. 2022. "Deepfakes and Society: What Lies Ahead?" In *Frontiers in Fake Media Creation and Detection*. Springer.

doi:10.1007/978-981-19-1524-6_1.

Hameleers, Michael. 2022. ""I Don't Believe ANything They Say Anymore!" Explaining Unanitcipated Media Effects Among Distrusting Citizens." *Media and Communication* 10 (3): 158-168.

doi:10.17645/mac.v10i3.5307.

Hameleers, Michael , Toni Van Der Meer, and Tom Dobber. 2022. "You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media." *Social Media + Society* 1-12. doi:10.1177/20563051221116346.

The Effect of Media Literacy on
Misinformation and Deep Fake Video Detection

Hilary, Ibegbulem Obioma, and Olannye-Okonofua Dumebi. 2021. "Social Media as a Tool for Misinformation and Disinformation Management." *Linguistics and Culture Review* 5 (S1): 496-505.

Hwang, Yoori , Ryu Ji Youn, and Se-Hoon Jeong. 2021. "Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education." *Cyberpsychology, Behavior, and Social Networking* 24 (3): 1-22. doi:10.1089/cyber.2020.0174.

Ivanov, Bobi, Kimberly A Parker, and Lindsay L LindsayDillingham. 2018. "Testing the Limits of Inoculation-Generated Resistance." *Western Journal of Communication* 82 (5): 648-665.

doi:10.1080/10570314.2018.1454600.

Jang, S Mo, and Joon K Km. 2018. "Third person effects of fake news: Fake news regulation and media literacy interventions." *Computers in Human Behavior* 80: 295-302. doi:10.1016/j.chb.2017.11.034.

Jones-Jang, S Mo, Tara Mortensen, and Jingjing Liu. 2019. "Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't." *American Behavioral Scientist* 65 (2): 371-388. doi:10.1177/0002764219869406.

Kietzmann, Jan, Adam J Mills, and Kirk Plangger. 2021. "Deepfakes: Perspectives on the Future "Reality" of Advertising and Branding." *International Journal of Advertising* 40 (3): 473-485.

doi:10.1080/02650487.2020.1834211

Iacobucci, Serena, Roberta De Cicco, Francesca Michetti, Riccardo Palumbo, and Stefano Pagliaro. 2021. "Deepfakes Unmasked: The Effects of Information Priming and Bullshit Receptivity on Deepfake Recognition and Sharing Intention." *CYBERPSYCHOLOGY, BEHAVIOR, AND SOCIAL NETWORKING* 24 (3): 194-202. doi:10.1089/cyber.2020.0149.

Mason, Lance E, Daniel G Krutka, and Jeremy Stoddard. 2018. "Media Literacy, Democracy, and the Challenge of Fake News." *Journal of Media Literacy Education* 10 (2): 1-10.

McDougall, Julian. 2019. "Media Literacy versus Fake News: Critical Thinking, Resilience and Civic Engagement." *Media Studies* 10 (19): 29-45. doi:10.20901/ms.10.19.2.

Neag, Annamária, Çiğdem Bozdağ, and Koen Leurs. 2022. "Media Literacy Education for Diverse Societies." In *Oxford Research Encyclopedia of Communication*, by S Roy, F Subervi and M Georgiou . Oxford: Oxford University Press. doi:10.1093/acrefore/9780190228613.013.1268.

Ng, Yu-Leung. 2022. "An Error Management Approach to Perceived Fakeness of Deepfakes: The Moderating Role of Perceived Deepfake Targeted Politicians' Personality Characteristics." *Current Psychology*. doi:0.1007/s12144-022-03621-x.

Nurcahyani, M T, F Junaedi, and E Sudiwijaya. 2022. *Digital Literacy: How Social Media Prevent Misinformation During Pandemic*. Vol. 1655, in *HCI International 2022 – Late Breaking Posters*, by C Stephanidis, M Antona, S Ntoa and G Salvendy. Cham: Springer. doi:10.1007/978-3-031-19682-9_8.

n.d. *Oxford Languages*. Definition of Detection. Accessed November 2, 2022.

n.d. *Oxford Languages*. Definition of Credibility. Accessed November 2, 2022.

Popescu, Maria Magdalena. 2020. "Media Literacy Tools in Combating Disinformation and Fake News on Social Media." *Bulletin of the Transilvania* 13 (62): 103-112. doi:10.31926/but.ssl.2020.13.62.1.11.

Severin, Werner Joseph, and James W Tankard. 2014. *Communication Theories: Origins, Methods, and Uses in the Mass Media*. Pearson.

Sharma, Karishma, Yizhou Zhang, and Yan Liu. 2022. "COVID-19 Vaccine Misinformation Campaigns and Social Media Narratives." *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media*. Atlanta: AAAI Press. 920-931. doi:10.1609/icwsm.v16i1.19346.

Shin, Soo Yun , and Jiyoung Lee. 2022. "The Effect of Deepfake Video on News Credibility and Corrective Influence of Cost-Based Knowledge about Deepfakes." *Digital Journalism* 10 (3): 412-432. doi:10.1080/21670811.2022.2026797.

Sikorski, Christian Von, Viorela Dan, Britt Paris, Joan Donovan, Michael Hamleers, Jon Roozenbeek, and Sander Van Der LInden. 2021. "Visual Mis- and Disinformation, Social Media, and Democracy." *Journalism & Mass Communication Quarterly* 98 (3): 641-664. doi:10.1177/10776990211035395.

Skafle, Ingjerd, Anders Nordahl-Hansen, Daniel S Quintana, Rolf Wynn, and Elia Gabarron. 2022. "Misinformation About COVID-19 Vaccines on Social Media: Rapid Review." *Journal of Medical Internet Research* 24 (8): e37367. doi:10.2196/37367.

Thaw, Nyein, Thin July, Aye Nu Wai, Dion Hoe-Lian Goh, and Alton Y. K. Chua. 2021. "How Are Deepfake Videos Detected? An Initial User Study." *23rd HCI International Conference*. Switzerland: Springer Nature. 631-636.

n.d. *The Britannica Dictionary*. Accessed November 2, 2022. https://www.britannica.com/dictionary/credibility.

Roozenbeek , Jon, and Sander van der Linden. 2021. *Inoculation Theory and Misinformation.* Latvia: NATO Strategic Communication Center of Excellence.

Westerlund, Mika. 2019. "The Emergence of Deepfake Technology: A Review." *Technology Innovation Management Review* 9 (11): 39-52.

doi:10.22215/timreview/1282.

Yildirim, Akin, and Ergün Yolcu. 2022. "How Deep is the Fake? Tracing the Concept of Deepfake Through Youtube." *Electronic Cumhuriyet Journal of Communication* 4 (1): 63-74. doi:10.54089/ecider.1110865.

Appendices

Appendix A:

<div align="center">The Fake Post of Mark Zukerburg</div>

Appendix B:

Link to the deepfake Mark Zukerburg video on YouTube

https://www.youtube.com/watch?v=cnUd0TpuoXI

<div align="center">**The Effect of Media Literacy on**
Misinformation and Deep Fake Video Detection</div>