A case study on Deepfake Awareness, Mumbai, India.

Mohak Dwarkadhish Sharma, Abdul Nasim Nadir Shaikh

1. (MS in Data Science, Analytics and Engineering, Arizona State University, United States)

2. (Masters in Computer Applications, Allana Institute of Management and Technology, India)

Abstract: *With the rapid growth of technology in both computing hardware and algorithms, Artificial Intelligence (AI) has potential in a wide range of fields such as face and speech recognition, image analysis, virtual assistance, search engines, and many more. Deepfake is one of those recently developed apps enabled by deep learning. Deepfake is a technique that people can use to make things appear to have happened in real life. Being simple to use, anyone without prior understanding of Deep Learning may create such media using numerous tools that are already on the market which may lead to misuse.*

In our study, we have created a questionnaire using HTML, CSS, and JavaScript, and Google Sheet. We studied this data to understand one's ability to identify deepfake among our participants. A total 91 of 190 participants (47.89%) have responded.

Our findings show that there was no significant association between education level in identifying Deepfake while, field and awareness have a significant association. ***Key word:*** *Deepfakes, Artificial Intelligence, Deep-learning, GANs, Education*

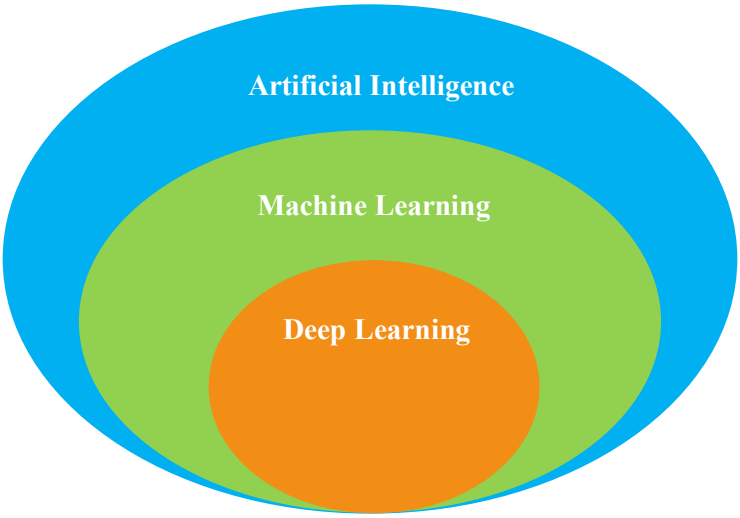
1.Introduction

Recent developments in Artificial Intelligence (AI) have had a significant impact on several fields, including common digital media, and have important ramifications. Due to the numerous real-world uses for AI that have been demonstrated in recent years, the majority of which are attributable to a discipline of AI called "Deep Learning", AI is seen as a paradigm-shifting technology. Deep learning is a subset of a more general field of AI called machine learning, which is predicated on the idea of learning from example. The more robust and complete the training data, the better the model gets. In deep learning, a model is able to automatically discover representations of features in the data that permit (1) classification or (2) parsing of the data. They are effectively trained at a "deeper" level.

DEEPFAKE, a term derived from the words "Deep Learning" and "fake content", is also called Synthetic reality. Deepfakes are made using methods that may overlay facial images of a target person onto videos of a source person in order to create videos of the target person acting or saying what the source person says. Face swap is a type of deepfake that includes this. Deepfakes, as defined more broadly, are pieces of entertainment created by AI that can also be categorized into two categories (1)lip-sync and (2) puppet-masters. A noteworthy cause for concern is the ease with which some people can be misled by fake videos and pictures that circulate online. Deepfake, which was named after a social media site Reddit account with the name "Deepfake" which then claimed to have developed a machine learning technique to transfer celebrity faces into adult content, is the practice of fabricating content by changing the face of a person serving as the source in an image or a video of another person serving as the target. Additionally, this method is employed to distribute hoaxes, commit fraud, and fabricate news stories.

Due to its simplicity and low cost of availability, Deepfake technology is becoming more and more popular today. Additionally, the variety of Deepfake techniques' applications sparks both expert and inexperienced users' curiosity. Deep autoencoders, which have two uniform deep belief networks with 4 or 5 layers each representing the encoding

half and the remaining layers the decoding half, are a popular deep network paradigm. Deep encoding is frequently used to compress and reduce the size of photographs. The first Deepfake approach success was the development of the FakeApp by the Reddit user "Deepfake" utilizing the autoencoder-decoder pairing structure, which is used to swap one person's face with another. For better outcomes, "FakeApp" software requires enormous amounts of data, which is provided to the system in order to train the model and then insert the source person's face into the desired video. To create false videos, all of the pictures from the source video must first be extracted into a folder, appropriately cropped, aligned, and processed using a trained model before the final film is ready.



Deepfake content might include pornography, politics, or bullying of a person by utilizing his or her image and voice without his or her consent. While traditional visual effects or computer graphics approaches can be used to create some deepfakes, the most recent common underlying mechanism for deepfake creation is deep learning models such as (1) autoencoders based on supervised learning models and (2) Generative adversarial networks (GANs) based on unsupervised learning models, which have been widely used in the computer vision domain.

2. Review of Literature

In the recent years, there have been a rapid increase in the use of deepfake technology. There have been many instances of such deepfake media circulating over the internet and social media platforms. Authors such as Dagar and Vishwakarma (2022)[1] in their article presents us with a survey on the application of deepfakes, followed by discussions on state-of-the-art methods for deepfake generation and detection for three media: image, video, and audio. It also discusses the architectural components and datasets used for various methods of deepfakes. A more extensive study by Thi-Nguyenaet.al, (2022)[2] accords us with a survey of algorithms used to create deepfakes and more importantly methods proposed to detect deepfakes in the literature to date.

Most of these study gives us an explanation on how to create and detect a deepfake using deeplearning techniques, but none of them discuss about the group of people who are affected by these deepfakes. If a group of people is already aware about these deepfakes, they are less prone to getting defrauded by someone. Also, researchers at the U.S. Army Combat Capabilities Development Command, known as DEVCOM, Army Research Laboratory, in collaboration with Professor C.-C. Jay Kuo’s research group at the California, developed a deepfake detection

method that will allow for the creation of state-of-the-art Soldier technology to support mission-essential tasks such as adversarial threat detection and recognition. The result is an innovative technological solution called DefakeHop (Chen, et al., 2021)[3].

3. Aims & Objectives

The objective of this study was to investigate the awareness and recognition of deepfake technology among the residents and mainly the students of a Mumbai (India) based local community.

4. Research Questions

The research questions are as follows:

1. To know how many participants in the local community are aware of the existence and implications of deepfake technology?
2. To know how many participants in the local community can correctly identify whether a given audiovisual media is real or fake?
3. Does having a technical background influence the awareness and recognition of deepfake technology?

Hypothesis

I. Users with IT background are aware more about Deepfake than those of non-IT background
II. Users with higher educational qualifications are more likely to detect the Deepfake comparatively than who are having lower education background

5. Methodology

To answer these questions, an online survey was conducted which was made available through our link .

A total 91 of 190 participants (47.89%) responded. The participants were invited from Mumbai (India) based local community. The community was selected based on convenience and accessibility. The participants were recruited through social media posts and word-of-mouth. The inclusion criteria were: having access to a smartphone or a computer with an internet connection. The exclusion criteria were: having a visual or hearing impairment that would prevent them from viewing or listening to the audiovisual media. The participants were informed about the purpose and procedures of the study, and they gave their informed consent before taking part in the survey.

Following were the study's materials:

- A questionnaire with two sections: demographic information and the identification of real and fake audiovisual media. The questionnaire was created using HTML, CSS, and JavaScript, and the data was saved in a Google Sheet for further analysis.
- A collection of ten audiovisual media clips, either real or fake. The clips were selected from various sources such as news outlets, social media platforms, and entertainment websites. The clips varied in length, quality, content, and context. Four clips were real, meaning that they were not manipulated or altered in any way.
- A scoring rubric was used to evaluate the participant’s responses to the recognition section of the questionnaire. The scoring rubric assigned one point for each correct answer (identifying a real clip as real or a fake clip as fake) and zero points for each incorrect answer (identifying a real clip as fake or a fake clip as real). The total score ranged from 0 to 10, with higher scores indicating better recognition of deepfake technology.

The procedure of this study was as follows:

- The participants received an invitation link to the questionnaire. The participants clicked on the link and accessed the online questionnaire.
- The participants then proceeded to fill out the questionnaire. They first answered some questions such as education level, and educational background. Then they answered a question about their awareness of deepfake technology, such as whether they had heard of it before. They then viewed 10 audiovisual media clips in random order and indicated whether they thought each clip was real or fake.
- The researcher collected and analyzed the data using descriptive and inferential statistics.
- Shapiro-Wilk tests value ($p<0.05$) indicate a non-normal distribution. Hence, non-parametric tests including Chi-square test, Mann-Whitney U, and Kruskal-Walli's test were employed.
- Mann-Whitney U tests were used for binary variables viz., (a) awareness and non-awareness, and (b) field IT and Non-IT.
- Kruskal-Walli's test was employed for the categorical variables the educational level and total scores
- The researcher calculated the frequency and percentage of participants who were aware of deepfake technology and compared them across different groups using chi-square tests of association and Fisher's exact test.

6. Results

A total of 91 records were taken into consideration for the study. The number of participants was then categorized into different groups according to their field, educational background, and awareness of the studied topic.

To determine the association among categorized variables viz., field (IT and Non-IT), education (high school, undergraduate, graduate and post graduate) and Total (Total marks) which shows
the ability to identify a media to be deepfake, chi-square test of association was employed.

Fisher's Exact test was also employed as more than 20% cell have value less than 5.

Table 1 First Summary

	Frequency	Percent
Background		
IT	35	38.5
Non-IT	56	61.5
Total	91	100
Education		
graduate	40	44
highschool	16	17.6
postGraduate	8	8.8
undergrad	27	29.7
Total	91	100

Awareness		
NO	48	52.7
YES	43	47.3
Total	91	100

Case 1: Field * Total

Fisher's Exact Test

The results of the Fisher's exact test (p = 0.025), which is less than our significance level of 0.05. Our results are statistically significant. We can reject the null hypothesis and conclude that a relationship exists between field and total marks.

It can be observed from Figure 1 (Box-Whisker Chart) , that there is a visible difference between the median values of both the fields. However, non-IT have few outliers.

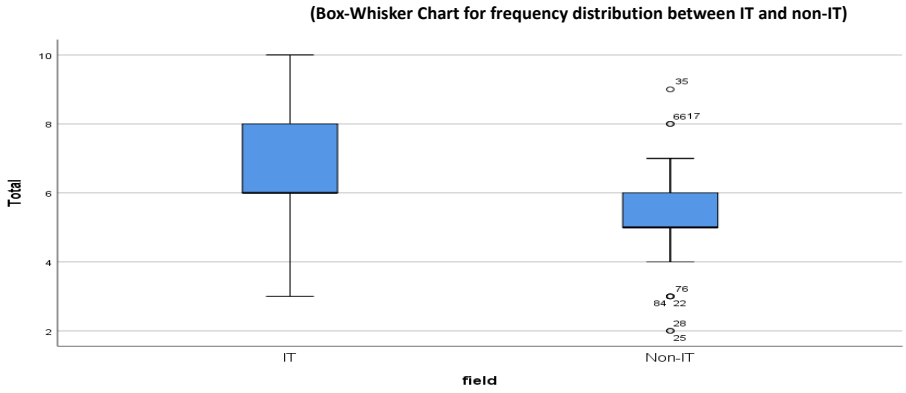


Figure 1

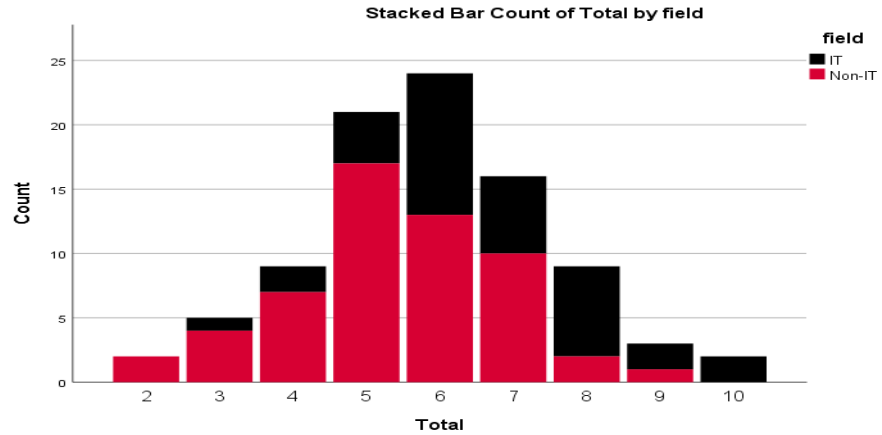
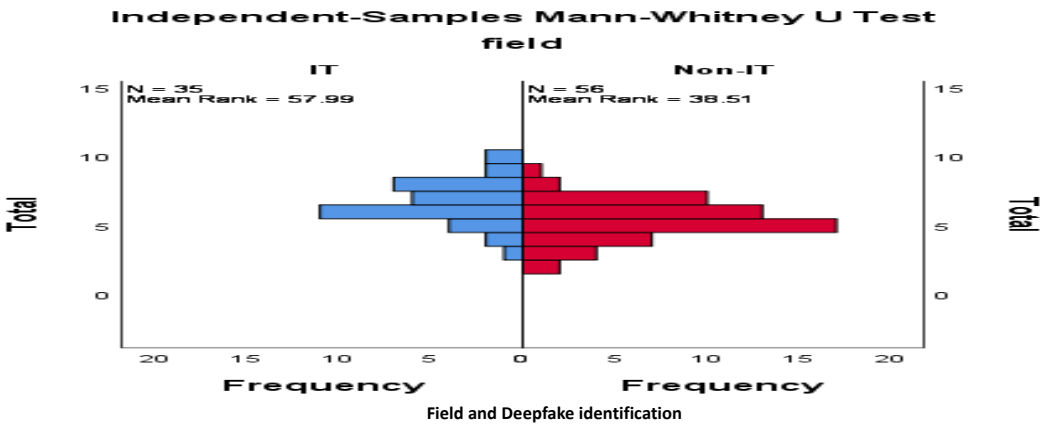


Figure 2

From Figure 2, it was found that the number of participants who could identify the media above the mean score (mean = 6) consists mainly of people from an IT background and vice versa.

Mann-Whitney U test (U = 560.5 , p = .000) shows that there is significant difference across the field (IT and Non-IT) and identifying deepfake.

Figure 3



Case 2: Education * Total

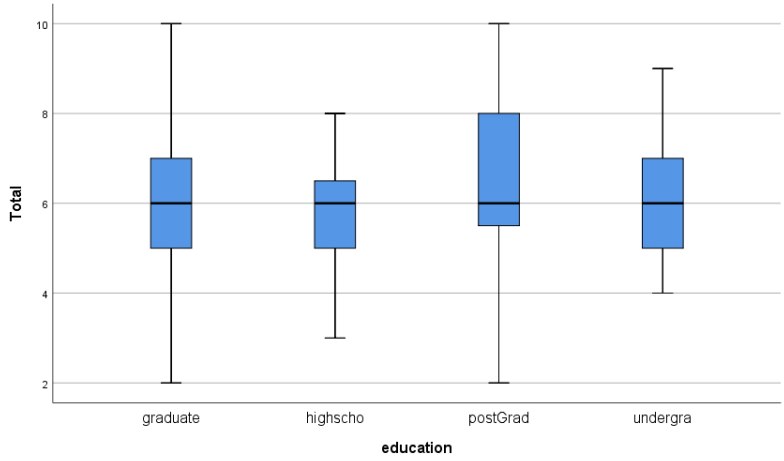


Figure 4

The results of Chi-square tests($\chi 2 = 23.684$,df =24, p>.05),hence the null hypothesis is accepted and it is concluded that there is no significant difference between the education level and one's ability to identify the deepfake.This results could be seen in the light of the fact that nowadays even many high school students are computer literatre. Though figure 4,depicts that high school category have values in the lower third quartile, while, post-graduate have majority of the values in the upper third quartile, this is not significant. Furthermore, Kruskal-Wallis test (test statistics = 1.032, df = 3, p = .794) shows that there is no significant difference among any pairs of the education level and identifying deepfake.

Case 3: Awareness * Total

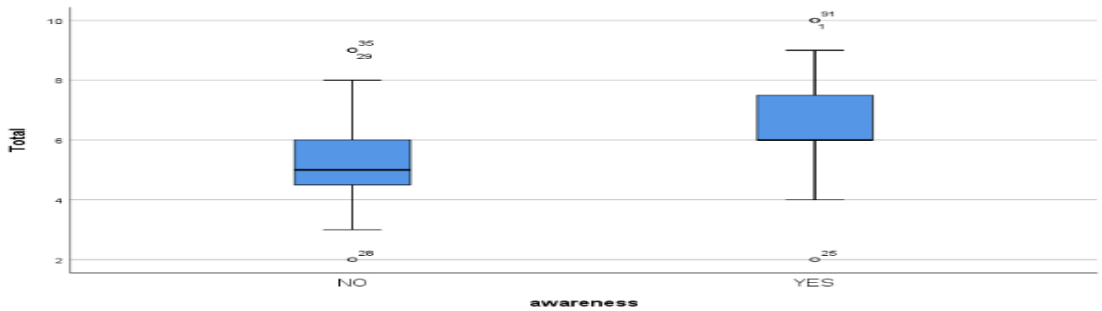


Figure 5

The Chi-Square test there was significant evidence of an association, ($\chi 2$ (8) = 20.828, $p < 0.05$) between the categorical variable's awareness and identification of deepfake. Since 12 cells (66.7%) have expected count less than 5Fisher's exact test was performed and p-values (0.003) was found, thus our results are statistically significant. Thus, we accepted the alternate hypothesis and conclude that a significant relationship exists between Awareness and total marks.

In addition, figure 5, shows that there is a visible difference between the median values of both the fields. Mann-Whitney U test (U = 1450 , p = .001) shows that there is significant difference across the awareness and identifying deepfake.

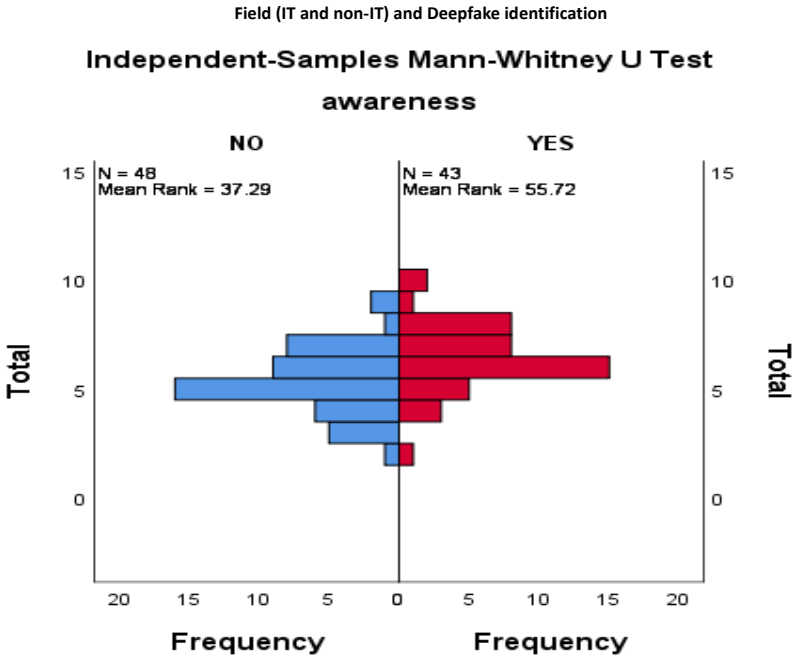


Figure 6

7. Conclusion

Our objective was to find the association between the education level, awareness, and field (technical or non-technical) in terms of identifying a media correctly as deepfake. In the first step we had applied Chi-Square test of association has shown a significant result for field, awareness however no significant association was seen for the education level. We further solidified this with Mann-Whitney U test and Kruskal-Wallis test.

Kruskal-Walli's test showed that there was no significant difference between education level and identification of deepfake. However, there was a significant difference between field, awareness and identification of media. The Mann-Whitney U test results indicated a significant difference in identifying deepfake those having IT than those with non-IT background.

References

1) Dagar, D., & Vishwakarma, D. K. (2022). A literature review and perspectives in deepfakes: generation, detection, and applications. International journal of multimedia information retrieval, 11(3), 219-289.

2) Nguyena, T. T., Viet, Q., Nguyenb, H., Nguyena, D. T., Nguyena, D. T., & Huynh-Thec, T. (2022). Deep Learning for Deepfakes Creation and Detection: A Survey. SSRN Electron. J, 223, 103525.

3) Chen, H. S., Rouhsedaghat, M., Ghani, H., Hu, S., You, S., & Kuo, C. C. J. (2021, July). Defakehop: A light-weight high-performance deepfake detector. In 2021 IEEE International conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.

4) Mahmud, B. U., & Sharmin, A. (2021). Deep insights of deepfake technology: A review. arXiv preprint arXiv:2105.00192.

5) Vyas, H. (2020). Deep fake creation by deep learning. Extraction, 7(07).