Journal of Economics, Finance and Management Studies

ISSN (print): 2644-0490, ISSN (online): 2644-0504

Volume 07 Issue 11 November 2024

Article DOI: 10.47191/jefms/v7-i11-34, Impact Factor: 8.044

Page No: 6910-6914

The Role of Whistleblowing in Unmasking the Phenomenon of Deepfake Fraud

Alim Dhisa Istiqomah[1], Makaryanawati[2], Diana Tien Irafahmi[3]

[1,2,3]Department of Accounting, Faculty of Economics and Business, Universitas Negeri Malang, Indonesia

ABSTRACT: Deepfake fraud is a form of threat that leverages Artificial Intelligence (AI) technology to gain financial or non-financial benefits by deceiving others. The protection mechanisms against deepfake fraud have yet to be regulated under clear legal frameworks. This study explores the role of whistleblowing as a method to reveal indications of deepfake fraud in organizational governance, given the limited coverage in prior literature. Utilizing a scoping review approach, the study examines data sources from Google Scholar and Scopus databases over a 10-year period (2014-2024). The research findings indicate that whistleblowing plays a role in addressing early-stage issues, raising individual awareness, shaping regulations related to deepfake fraud, protecting organizational integrity, and maintaining public trust. Reporting instances of deepfake fraud promptly helps organizations and individuals safeguard vital assets or data security.

KEYWORDS: Artificial Intelligence, Deepfake Fraud, Financial Fraud, Non-Financial Fraud, Whistleblowing

I. INTRODUCTION

Fraud that leverages the role of technology is increasingly prevalent in various countries around the world. This digital crime can manipulate its targets into believing the authenticity of fabricated images, videos, and audio (Bateman, 2020; Cross, 2022) Support from multiple parties is crucial in uncovering the phenomenon of deepfake fraud, one of which is through whistleblowing. This is because 43% of the fraud reports received by the ACFE in 2024 came from whistleblowing. The role of whistleblowing has provided significant benefits for organizations in protecting their assets from fraud (Dávila et al., 2023). Whistleblowing can be conducted internally within a company or externally, thus making it easier for witnesses to report misconduct (Dworkin & Baucus, 1998; Latan et al., 2021).

The use of Artificial Intelligence (AI) has shifted the role of technology towards negative applications. The efficiency of AI in managing information is being exploited as a medium for fraud. Deepfake technology, stemming from the ease provided by AI, uses collections of image, video, or audio sources to manipulate information (Firc et al., 2023). Consequently, deepfakes have become a real threat to both organizations and individuals as a medium for fraud ((de Rancourt-Raymond & Smaili, 2023). The impact of deepfake fraud is highly complex, potentially leading to significant financial losses for its victims (Bateman, 2020). The misuse of deepfakes also jeopardizes the security of essential data for individuals and certain communities (de Rancourt-Raymond & Smaili, 2023).

Deepfake fraud is predicted to become the largest source of fraud targeting the assets of organizations or individuals in the coming years if preventive measures are not maximized (Trend Micro, 2019). Generally, deepfake fraud can cause business instability due to security issues that make organizations vulnerable to crime (Medius, 2024). A survey by PWC in 2022 on fraud revealed that the highest cases of financial fraud are not conducted traditionally, but rather innovatively by leveraging technology. One notable case of deepfake fraud occurred at a company in Hong Kong, where it targeted financial employees by faking the voice and video of the company's executive, resulting in a loss of up to $25 million ((Aprilia, 2024; Chen & Magramo, 2024; ET Online, 2024; Milmo, 2024).

The increasing distribution of deepfake fraud content targeting both organizations and individuals requires collective vigilance in identifying signs of digital fraud. Based on this background, this study aims to analyze the role of whistleblowing in uncovering the phenomenon of deepfake fraud. This study differs from previous research as it focuses on exploring whistleblowing as a risk management effort and the dangers of deepfake fraud content in the accounting, finance, and taxation sectors. This approach is taken considering that deepfake fraud often targets the financial sector. As a result, detrimental financial or non-financial impacts on organizations and individuals can be avoided.

II. LITERATURE REVIEW

A. Concept of Whistleblowing

Whistleblowing refers to the reporting of unethical or illegal actions that involve abuses of power, fraud, and other violations *(Lam & Harcourt, 2019)*. A whistleblower can report violations within their employing organization or to external (Alleyne et al., 2017; Soni et al., 2015). The whistleblowing process considers factors such as effectiveness, confidentiality, safety, and potential impact, as whistleblowers often face social backlash that can affect their mental and physical well-being. The motivation for whistleblowing varies depending on individual objectives. According to *Dozier & Miceli (1985) and Near & Miceli (1995)* individuals may engage in whistleblowing due to a desire to address injustices, protect public interests, or fulfill personal motivations. Whistleblowing has a significant impact on promoting transparency and accountability. Consequently, the reports provided can serve as valuable input in fostering more integrity-driven management.

B. Concept of Deepfake Fraud

Artificial Intelligence (AI) has developed a technology known as deepfake. Deepfake technology produces realistic but fake images, videos, and audio (Cross, 2022; Lu & Chu, 2023; Maras & Alexandrou, 2019). Initially, deepfake technology was beneficial for industries focused on fashion, film, and even gaming (de Rancourt-Raymond & Smaili, 2023). As it allowed for the rapid creation of thousands of content pieces. However, the development of deepfake technology has now shifted towards being exploited to create fraudulent content. Many irresponsible parties use deepfakes to impersonate others for financial gain (Levine, 2020).

Deepfake fraud has emerged as a new term to describe the negative applications of this technology. The concept of deepfake fraud involves schemes such as sending messages, making phone calls, or emailing organizations or individuals (Wilder, 2020). The perpetrators of deepfake fraud go through several stages before creating fraudulent content, such as gathering information about the target, identifying the person to impersonate, and then executing the scheme on the target (Firc et al., 2023). The impact of deepfake fraud is highly complex, yet its mitigation remains limited due to the lack of specific legal regulations (van der Sloot & Wagensveld, 2022). If deepfake fraud targets a company, it may be aimed at financial exploitation or damaging the organization's reputation.

III. METHOD

This study employs a literature review approach using a scoping review method. A literature review was chosen to systematically discover, classify, identify, evaluate, and interpret previous research on the topic (Creswell & Poth, 2016; Snyder, 2019). The literature review method using a scoping review is particularly useful when the research topic is new, complex, or has not been extensively studied (Peters et al., 2021). The goal is to gain a comprehensive understanding of "The Role of Whistleblowing in Uncovering the Phenomenon of Deepfake Fraud." The data sources used in this study include articles, books, conference proceedings, and research findings published in reputable databases such as Google Scholar and Scopus. The digital or online-based data search expands the range of sources and improves time efficiency. The literature search focuses on articles published within the 10-year period from 2014 to 2024 to ensure data relevance. All studies related to whistleblowing and deepfake fraud are observed without imposing criteria on language, type of publication, or specific geographical areas. The data collection process follows these stages (Arksey & O'Malley, 2005):

1. Identifying the research questions;
2. Identifying studies relevant to the topic;
3. Selecting the studies;
4. Mapping the data;
5. Interpreting the results.

The identified literature is managed using Mendeley as the reference management software. The validity of this study is ensured through discussions with experts to minimize bias in the selection of articles.

IV. RESULT AND DISSCUSSION

Based on the literature review conducted using reputable data sources, it has been found that employee whistleblowing can serve to protect organizations from the dangers of deepfake fraud. Whistleblowing has become one of the actions used to expose deepfake fraud due to the lack of regulations on prevention procedures or follow-up actions after such incidents (Chandra & Snowe, 2020). In fact, information from deepfake fraud has negative implications for various stakeholders, including individuals and organizations (Heidari et al., 2024; Muhammad & Hossain, 2022). Therefore, exposing deepfake fraud requires support from society, private organizations, and the government. This literature review will explore the discussion of deepfake fraud disclosure through the role of whistleblowing. This is due to the complex nature and difficulty in detecting deepfake fraud, which results in whistleblowing playing a key role in processing findings, as will be discussed in detail in the following section:

A. Whistleblowing Solves Initial-Level Issues

The implementation of Artificial Intelligence (AI) through the development of deepfakes has been widely utilized as a tool for fraud (Zhang et al., 2023). Human involvement is necessary as a preventive measure against deepfake fraud, such as through whistleblowing actions (Miller, 2023). The goal is for these reports to minimize the occurrence of illegal acts and public harm (Lam & Harcourt, 2019). Whistleblowing plays a key role as the first problem-solver that helps identify deepfake fraud cases. A whistleblower can provide relevant evidence to uncover digital manipulation, such as fake videos or audio, thereby assisting in further investigations. This is due to the increasing number of deepfake fraud cases, such as unusual video or audio recordings targeting individuals and causing significant financial harm. Therefore, individual reporting of anomalies or suspicious activities through whistleblowing channels helps organizations find solutions quickly. Thus, significant impact of deepfake fraud in the accounting, finance, and taxation sectors does not pose a sustained risk.

B. Whistleblowing Increases Individual Awareness of Deepfake Fraud

The role of whistleblowing in exposing deepfake fraud also depends on employees' awareness and knowledge of this phenomenon. Organizations need to provide training for employees, especially those working in accounting, finance, or tax sectors, on how to identify, prevent, and handle deepfake fraud. In the digital era, the workforce is expected to have an understanding of technology to ensure effective and efficient organizational governance (Blau et al., 2020). One of the benefits of having strong technological skills is that individuals can help identify deepfake violations quickly (Losbichler & Lehner, 2021). Active participation in learning about technological advancements can increase the likelihood of reporting deepfake fraud (Deloitte, 2024). This is because individuals often have direct access to the source of the issue, so internal information helps identify not only the forged content but also those involved in its creation or distribution methods. The identification of the whistleblower can ultimately assist authorities in accelerating the investigation, prosecution, or other legal actions.

C. Whistleblowing Shapes Regulatory Protections Against Deepfake Fraud

Whistleblowing helps organizations establish clear regulations to maximize the effectiveness of handling deepfake fraud. This policy can be formulated based on whistleblower information to ensure that the actions taken are well-targeted. Furthermore, the regulations must include protections for whistleblowers, secure and anonymous reporting channels, and transparent procedures for following up on reports. This is particularly important as deepfake fraud control regulations are limited in countries such as Singapore, China, Australia, Finland, Japan, Canada, the United States, the United Kingdom, and India (Cumming et al., 2024). According to Birrer & Just (2024), whistleblowers often face negative consequences, such as threats or discrimination from deepfake fraud developers. Some of these possibilities could occur, such as a decrease in an individual's willingness to report, which results in a lower level of control function. Therefore, policies regarding deepfake fraud need to be strictly regulated, taking into account whistleblower reports, so that individuals feel safer when reporting deepfake fraud findings. These regulations also reduce the risk of inaccurate reports and ensure that complaints are addressed promptly.

D. Whistleblowing Protects Organizational Integrity and Public Trust

According to Bateman (2020), the phenomenon of deepfake indirectly has the potential to damage public trust in organizations, such as by lowering their reputation. Public perceptions of data breaches and the loss of assets can also decrease a company's value (de Rancourt-Raymond & Smaili, 2023). Given the broad impact of deepfake fraud, whistleblowing is essential to create pressure on individuals or organizations to take responsibility for fostering a more ethical environment. The establishment of a whistleblowing system allows organizations to gather various sources of information related to deepfakes, enabling them to plan long-term risk management strategies for the technology.

V. CONCLUSIONS

This literature review aims to explore the important role of whistleblowing in uncovering the phenomenon of deepfake fraud in society, particularly in vulnerable sectors such as accounting, finance, and taxation. The findings of this study suggest that whistleblowing helps address early-stage issues, raises individual awareness, and shapes regulations related to deepfake fraud. Whistleblowing is an essential tool in protecting organizations from the threat of deepfake fraud. The results of this study indicate that whistleblowing helps address initial-level problems, raises individual awareness, shapes regulations related to deepfake fraud, protects organizational integrity, and fosters public trust. However, the results of this exploration using a literature review approach are still limited, as the accuracy of its implementation has not been directly proven. Therefore, future researchers may consider employing empirical studies. It is hoped that empirical studies and literature reviews will broaden the investigation into the role of whistleblowing in exposing deepfake fraud, which can help mitigate the risks of both financial and non-financial losses.

REFERENCES

1) ACFE. (2024). *Occupational Fraud 2024: A Report to the Nations*.
2) Alleyne, P., Charles-Soverall, W., Broome, T., & Pierce, A. (2017). Perceptions, predictors and consequences of whistleblowing among accounting employees in Barbados. *Meditari Accountancy Research*, *25*(2), 241–267. https://doi.org/10.1108/MEDAR-09-2016-0080
3) Aprilia, Z. (2024). *Pekerja Keuangan Ini Kena Tipu Rp392 M, Pelaku Pakai Deepfake*. CNBC Indonesia. https://www.cnbcindonesia.com/market/20240205155021-17-512018/pekerja-keuangan-ini-kena-tipu-rp392-m-pelaku-pakai-deepfake
4) Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology: Theory and Practice*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616
5) Bateman, J. (2020). *Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios Cyber Policy Initiative Working Paper Series | "Cybersecurity and the Financial System" #7*.
6) Birrer, A., & Just, N. (2024). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media and Society*. https://doi.org/10.1177/14614448241253138
7) Blau, I., Shamir-Inbal, T., & Avdiel, O. (2020). How does the pedagogical design of a technology-enhanced collaborative academic course promote digital literacies, self-regulation, and perceived learning of students? *Internet and Higher Education*, *45*. https://doi.org/10.1016/j.iheduc.2019.100722
8) Chandra, A., & Snowe, M. J. (2020). A taxonomy of cybercrime: Theory and design. *International Journal of Accounting Information Systems*, *38*. https://doi.org/10.1016/j.accinf.2020.100467
9) Chen, H., & Magramo, K. (2024). *Finance worker pays out $25 million after video call with deepfake 'chief financial officer' | CNN*. CNN. https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html
10) Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
11) Cross, C. (2022). Using artificial intelligence (AI) and deepfakes to deceive victims: the need to rethink current romance fraud prevention messaging. *Crime Prevention and Community Safety*, *24*(1), 30–41. https://doi.org/10.1057/s41300-021-00134-w
12) Cumming, D., Saurabh, K., Rani, N., & Upadhyay, P. (2024). Towards AI ethics-led sustainability frameworks and toolkits: Review and research agenda. *Journal of Sustainable Finance and Accounting*, *1*, 100003. https://doi.org/10.1016/j.josfa.2024.100003
13) Dávila, A., Derchi, G. B., Oyon, D., & Schnegg, M. (2023). External complexity and the design of management control systems: a case study. *Management Accounting Research*. https://doi.org/10.1016/j.mar.2023.100875
14) de Rancourt-Raymond, A., & Smaili, N. (2023). The unethical use of deepfakes. *Journal of Financial Crime*, *30*(4), 1066–1077. https://doi.org/10.1108/JFC-04-2022-0090
15) Deloitte. (2024, September 17). *Half of Executives Expect More Deepfake Attacks on Financial and Accounting Data in Year Ahead – Press release | Deloitte US*. Deloitte. https://www2.deloitte.com/us/en/pages/about-deloitte/articles/press-releases/deepfake-attacks-on-financial-and-accounting-data-rising.html
16) Dozier, J. B., & Miceli, M. P. (1985). Potential predictors of whistle-blowing: A prosocial behavior perspective. *Academy of Management Review*, *10*(4), 823–836.
17) Dworkin, T. M., & Baucus, M. S. (1998). Internal vs. External Whistleblowers: A Comparison of Whistleblowering Processes. *Journal of Business Ethics*, *17*, 1281–1298. https://doi.org/https://doi.org/10.1023/A:1005916210589
18) ET Online. (2024). *Hong Kong MNC suffers $25.6 million loss in deepfake scam - The Economic Times*. The Economic Times. https://economictimes.indiatimes.com/industry/tech/hong-kong-mnc-suffers-25-6-million-loss-in-deepfake-scam/articleshow/107465111.cms?from=mdr
19) Firc, A., Malinka, K., & Hanáček, P. (2023). Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors. In *Heliyon* (Vol. 9, Issue 4). Elsevier Ltd. https://doi.org/10.1016/j.heliyon.2023.e15090
20) Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. (2024). Deepfake detection using deep learning methods: A systematic and comprehensive review. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 14, Issue 2). John Wiley and Sons Inc. https://doi.org/10.1002/widm.1520

21) Lam, H., & Harcourt, M. (2019). Whistle-blowing in the digital era: motives, issues and recommendations. *New Technology, Work and Employment*, *34*(2), 174–190. https://doi.org/10.1111/ntwe.12139

22) Latan, H., Chiappetta Jabbour, C. J., & Lopes de Sousa Jabbour, A. B. (2021). Social Media as a Form of Virtual Whistleblowing: Empirical Evidence for Elements of the Diamond Model. *Journal of Business Ethics*, *174*(3), 529–548. https://doi.org/10.1007/s10551-020-04598-y

23) Levine, A. J. (2020). *Dollars, Deception, and Deepfakes: An Analysis of Deepfakes and Synthetic Media Fraud*.

24) Losbichler, H., & Lehner, O. M. (2021). Limits of artificial intelligence in controlling and the ways forward: a call for future accounting research. *Journal of Applied Accounting Research*, *22*(2), 365–382. https://doi.org/10.1108/JAAR-10-2020-0207

25) Lu, H., & Chu, H. (2023). Let the dead talk: How deepfake resurrection narratives influence audience response in prosocial contexts. *Computers in Human Behavior*, *145*. https://doi.org/10.1016/j.chb.2023.107761

26) Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence and Proof*, *23*(3), 255–262. https://doi.org/10.1177/1365712718807226

27) Medius. (2024). *An Accounting of Financial Professionals*.

28) Miller, M. (2023). *Deepfakes: Real threat*.

29) Milmo, D. (2024). *Company worker in Hong Kong pays out £20m in deepfake video call scam | Hong Kong | The Guardian*. TheGuardian.https://www.theguardian.com/world/2024/feb/05/hong-kong-company-deepfake-video-conference-call-scam

30) Muhammad, G., & Hossain, M. S. (2022). Light deep models for cognitive computing in intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, *24*(1), 1144–1152.

31) Near, J. P., & Miceli, M. P. (1995). Effective-whistle blowing. *Academy of Management Review*, *20*(3), 679–708.

32) Peters, M. D. J., Marnie, C., Colquhoun, H., Garritty, C. M., Hempel, S., Horsley, T., Langlois, E. V., Lillie, E., O'Brien, K. K., Tunçalp, Özge, Wilson, M. G., Zarin, W., & Tricco, A. C. (2021). Scoping reviews: reinforcing and advancing the methodology and application. In *Systematic Reviews* (Vol. 10, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13643-021-01821-3

33) PWC. (2022). *PwC's Global Economic Crime and Fraud Survey 2022*.

34) Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339.

35) Soni, F., Maroun, W., & Padia, N. (2015). Perceptions of justice as a catalyst for whistle-blowing by trainee auditors in South Africa. *Meditari Accountancy Research*, *23*(1), 118–140. https://doi.org/10.1108/MEDAR-01-2014-0004

36) Trend Micro. (2019). *Trend Micro Security Predictions for 2020*. https://documents.trendmicro.com/assets/rpt/rpt-the-new-norm-trend-micro-security-predictions-for-2020.pdf

37) van der Sloot, B., & Wagensveld, Y. (2022). Deepfakes: regulatory challenges for the synthetic society. *Computer Law and Security Review*, *46*. https://doi.org/10.1016/j.clsr.2022.105716

38) Wilder, M. (2020). *Fooling the senses for profit*. Fraud Magazine. https://www.fraud-magazine.com/article.aspx?id=4295009426

39) Zhang, C., Zhu, W., Dai, J., Wu, Y., & Chen, X. (2023). Ethical impact of artificial intelligence in managerial accounting. *International Journal of Accounting Information Systems*, *49*. https://doi.org/10.1016/j.accinf.2023.100619