



Issues in Informing Science + Information Technology

An Official Publication
of the Informing Science Institute
InformingScience.org

IISIT.org

Volume 22, 2025

THE PREDOMINANT ETHICAL ISSUES AROUND DEEP FAKE TECHNOLOGY AND FAKE NEWS ON SOCIAL MEDIA

Singarila Somanje	University of the Witwatersrand, Johannesburg, South Africa	Singarila.somanje@outlook.com
John Mangundu*	University of the Witwatersrand, Johannesburg, South Africa	john.mangundu@wits.ac.za

* Corresponding author

ABSTRACT

Aim/Purpose	This paper seeks to unearth the benefits of deep fake technology and its potential for application to pursue unethical intentions on social media, thereby negatively impacting individuals and society's well-being.
Background	The research paper addresses the problem by exploring the ethical implications of deep fake technology and fake news on social media. Through the analysis of the impact on trust, privacy, and democracy, regulatory and accountability recommendations are made.
Methodology	Through a systematic literature review and thematic data analysis, this paper presents interesting ethical issues around deep fake technology and fake news on social media.
Contribution	This study contributes to the debate around artificial intelligence, social media, and the associated regulatory environment by offering insights into deep fake technology's social, political, and psychological consequences.
Findings	The study finds an urgent need to design and implement a strong regulatory framework for both content creators and social media platforms to curb the spread of harmful content and protect individuals' rights.
Recommendations for Practitioners	The study recommends robust content moderation, stronger regulatory frameworks, media literacy, and awareness campaigns to citizens to improve their ability to assess social media content's authenticity.
Recommendations for Researchers	Researchers are encouraged to take an interdisciplinary approach that includes law, ethics, information systems, psychology, and media studies to address the challenges brought by deep fake technology.

Accepting Editor: Eli Cohen | Received: October 29, 2024 | Revised: February 21, 2025 |

Accepted: April 25, 2025

Cite as: Somanje, S., & Mangundu, J (2025). The predominant ethical issues around deep fake technology and fake news on social media. *Issues in Informing Science and Information Technology*, 22, Article 2.

<https://doi.org/10.28945/5504>

(CC BY-NC 4.0) This article is licensed to you under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

Impact on Society	The paper impacts society by advancing the comprehension of the potential impacts of digital manipulation.
Future Research	Scholars may conduct longitudinal studies to determine the long-term psychological and social effects of individuals' exposure to deep fakes on social media.
Keywords	social media, ethics, deep fake technology, fake news, artificial intelligence, deep learning

INTRODUCTION

The use of social media in the 21st century is growing at a rapid rate (Barrett-Maitland & Lynch, 2020). By 2025, there are predictions that there will be over 4.4 billion monthly active social media users (Statista, 2024). The primary benefit of social media use is that it breaks the distance barrier between people when it comes to information sharing. Social media provides accessible platforms to stay connected with friends and family, stay updated on the news, enables businesses to communicate with customers, and prompts users to engage in discussions to exchange ideas or thoughts on particular topics (Barrett-Maitland & Lynch, 2020; Dhiman, 2023; Zhou & Zafarani, 2020). However, for all of its benefits and uses, social media has the potential to fuel the dissemination of fake news and misinformation, which can seriously impact society (Botha & Pieterse, 2020; Dhiman, 2023). In today's era, fake news has been accepted as a global pandemic (Botha & Pieterse, 2020). For example, Dhiman (2023) and Botha and Pieterse (2020) reported that social media can give individuals or people of power (e.g., politicians and leaders) the ability to influence people's actions or decisions by giving out misinformation or disinformation. Fake news can negatively affect individuals as it persuades them to accept false information or biased stories as the truth.

In light of the fake news and misinformation discussed above, an emerging technology called a "deep fake" is portrayed as a new form of fake news in society (Botha & Pieterse, 2020; Brooks, 2021). Allcott and Gentzkow (2017) and Shu et al. (2020) define fake news as news that intends to deceive and mislead people's perceptions. Furthermore, Lazer et al. (2018) state that fake news is a subcategory of misinformation. Fake news can be portrayed as an image, article, message, media, story, or news (Meel & Vishwakarma, 2020). On the contrary, a deep fake is an amalgamation portmanteau of deep learning and "fake imagery" (Wagner & Blewer, 2019). A deep fake is a form of fake news. It is a manipulation technique that incorporates deep learning, a subsidiary of artificial intelligence, to create fake videos, audio recordings, or images by swapping an individual's face with another individual; usually, they use a well-known person. Deep fake technology is created using deep learning, which allows the computer to keep generating photos by learning from previous images; thus, it gets better and better at creating an accurate description of a deep fake, which can lead to the creation of new content over time (Albahar & Almalki, 2019; Korshunov & Marcel, 2018; Tolosana et al., 2020; Wagner & Blewer, 2019). According to Harris (2018), a deep fake is used to either blackmail or take revenge; for example, uploading a pornographic video of someone in order to humiliate or threaten a person (Meskys et al., 2020).

Deep fakes as a form of fake news have become increasingly popular due to their ability to create hyper-realistic outputs from the data collected, primarily photos and videos. Wagner and Blewer (2019) state that any layperson (someone with no technology background) can create a deep fake as it is widely available and easily accessible to the general public (Botha & Pieterse, 2020). Meskys et al. (2020) and Gamage et al. (2022) state that deep fake technology is rapidly improving, and it has raised several ethical concerns as it makes it difficult to draw the lines between what is considered accurate and what is considered fake. Deep fake technology can potentially deceive people and have an impact on people's private and social lives, as it can provide the tool to tamper with evidence in court cases, create revenge pornography, or misrepresent particular groups of people (Brooks, 2021; de Ruiter,

2021; Gamage et al., 2022). The first deep fake ever created was a pornographic video of famous actors (Gal Gadot, Emma Watson, Scarlet Johansson) whose faces were put on pornographic actresses, which was created by a Reddit user named “deep fake” (Brooks, 2021; Gamage et al., 2022; Pérez Dasilva et al., 2021; Wagner & Blewer, 2019; Wahl-Jorgensen & Carlson, 2021).

Interestingly, deep fake technology is quite a controversial topic (Gamage et al., 2022). While deep fake technology undermines the ability to trust content or information, the same technology has the potential to open a wide range of creative opportunities. Worryingly, deep fakes have already negatively impacted society, thus raising several ethical concerns with this technology. For example, Wagner and Blewer (2019) and Meskys et al. (2020) identify some ethical concerns, such as invasion of privacy emanating from nonconsensual graphic content. Human deception is another ethical concern identified, as it is becoming increasingly difficult to detect the truth from images, videos, and news, leading to people having false memories (Brooks, 2021; de Ruiter, 2021; Liv & Greenbaum, 2020; Wagner & Blewer, 2019). Online harassment, cyberbullying, spreading fake news, misinformation, and issues of consent are other ethical concerns that are up for discussion (Gamage et al., 2022; Meskys et al., 2020). In the same context, Brooks (2021) states that deep fake technology can have the power to influence many different sectors, such as politics, businesses, the military, and national security.

Liv and Greenbaum (2020) and Smith and Mansted (2020) further emphasize that deep fakes have the potential to ruin reputations, create impersonations, and blackmail people by influencing their perception of reality. On a broader scale, this can negatively affect democracies as they can misrepresent specific individuals or nations, portray fake events such as terrorist attacks, and may deepen the division amongst social groups. Furthermore, deep fakes can create new tools for cybercrime operations, produce more online propaganda disinformation, and decrease trust in certain institutions (Meskys et al., 2020; Smith & Mansted, 2020). Deep fakes are an issue as they further enhance fake news, which harms people’s perception of reality and the truthfulness of information (de Ruiter, 2021; Meskys et al., 2020).

Any individual can become a victim of deep fake technology because it has the potential to damage one’s life permanently. After all, once the video is on social media platforms, it will travel quickly; thus, the damage can be irreversible even if the post is removed (de Ruiter, 2021; Helmus, 2022; Lazer et al., 2017). Deep fake technology is not something the law can protect one from, as the people who release these harmful videos or audio are not being held accountable; hence, it is important to discuss its impact on individuals and society so that solutions can be made to combat the risks of such technologies and lessen the harm it can cause on ordinary human beings (Chesney & Citron, 2019).

Despite substantial literature on what deep fakes and fake news are, how they are detected, and the types of detection techniques (Farid, 2022; Koopman et al., 2018; Korshunov & Marcel, 2018; Liang et al., 2023; Tolosana et al., 2020; Yang et al., 2019; Zhou & Zafarani, 2020; Zhou et al., 2019), there is little research on the associated risks brought by deep fake technologies and the impact of such risks on individuals or social media users (Helmus, 2022; Lazer et al., 2018; Vaccari & Chadwick, 2020). Therefore, this paper aims to contribute to scholarly discussions around deep fakes by answering the following research questions:

RESEARCH QUESTIONS

1. What are the types of fake news?
2. How does the use of deep fakes benefit and challenge the digital society?
3. What are the predominant ethical issues around deep fake technology and fake news on social media?
4. What are the societal implications of using deep fake technology?

RESEARCH METHOD

DATA SOURCES AND SEARCH TERMS

To answer the research questions and achieve the research objective, a comprehensive systematic literature search was conducted over eight databases: EBSCOhost, Google Scholar, ProQuest, Science Direct, Web of Science, Taylor and Francis, Research Gate, and JSTOR. These databases were selected as they include a variety of journals from various fields of study and are popular amongst researchers. Various search terms were used to find the articles needed for the systematic review between June and September 2023. Each search item had specific wildcards and quotation marks to yield the most accurate search results. Each search would use filters including all search terms from abstracts, titles, or anywhere in the article. The search terms included were fake news or false news, misinformation, deep fakes, ethics or ethical concerns, and ethical issues (Table 1). This paper was constructed using the PEO framework. This framework will assist in assessing the ethical issues arising in society after exposure to deep fake technology and the implications it can cause to social media or internet users. This framework can illustrate the aftermath effect that happens to individuals exposed to deep fake technology and its impact on society.

Furthermore, the PEO framework helped justify these ethical concerns that were identified as being related to deep fake technology. The people of interest (**P**) in this review are social media users, internet users, or individuals who use the internet. Exposure (**E**) refers to deep fake technology, or a form of fake news, and (**O**) is the ethical implications identified from using this technology and how it contributes as a new form of fake news. By searching the various databases to locate the relevant literature, the following keywords relate to the research question: *Systematic Review, Ethics, Deep Fakes, Fake News, Artificial Intelligence, and Deep Learning*. Table 1 lists all the search queries conducted and their database.

Table 1. Database and search terms

	Database name	Search query
1.	EBSCOhost, ProQuest, JSTOR, Springer Link, Taylor and Francis, Science Direct, Web of Science	“Ethical Issues” AND “Deep Fakes” “Ethical Issues” AND “Deep Fakes” NOT Covid “Deep fakes” NOT “Fake news” “Deep fakes” AND “Fake news” (“Deep Fakes” AND “users”) AND la:(eng OR en) NOT “covid - 19” “Society” AND “Deep Fakes” forms of fake news
2.	Google Scholar, Research Gate	allintitle: deepfakes ethics “Deepfake *” AND “ethical issues” (Deepfakes) on (social) AND (psychological) Fake news AND (ethics OR ethical issues) Different types of fake news

INCLUSION AND EXCLUSION CRITERIA

The paper considers articles, reports, and books published between 2017 and 2023, written in English. The reason for these criteria is that the first deep fake video was published in 2017. Hence, more literature started to develop from this time till the present, 2023. Importantly, scholarly material mainly focuses on the ethical issues of deep fakes and the implications of deep fake news as a form of fake news. However, non-English articles, magazines, and letters were excluded. In addition, non-

full-text and non-open-access articles were excluded. Lastly, articles focusing on fake news in the context of COVID-19 were excluded as they were out of the study scope.

BIAS

The following biases were assessed in this systematic literature review and how they were addressed. The first bias pertained to reporting bias, when a researcher puts or allocates search terms to yield positive results. The researchers addressed this bias by adhering to the research protocol and formulating a central research question with sub-questions (Drucker et al., 2016). Evidence selection bias is when the researcher has not studied the entire population of interest to know what should be included in the selection of articles, which leads to bias in the review. The researchers addressed this bias by reviewing articles on our research topic, meaning that all aspects have been thoroughly read and assessed. By clearly describing the topic of interest, we have reduced the selection bias (Drucker et al., 2016). Location bias occurs when one refers to a researcher only reviewing one type of database or when the relevant material for the review is only located in one area. The researchers addressed this bias by finding articles in different databases to yield various search results and by implementing searches on well-known databases to include all possible results (Drucker et al., 2016).

DATA EXTRACTION

In this systematic review, a data extraction sheet was used to record data obtained and searches conducted on databases. This data extraction sheet served as a guideline for how the literature would be used to answer the main and subquestions for this systematic literature review. The overall search results identified for this study were a total of 15,759. Fourteen duplicates were found on the databases and were removed, thus leaving 15,745. Of these articles, 15,471 were excluded after reading the titles as they met the exclusion criteria and did not relate to the research questions. After the excluded articles were removed, 288 articles were found. After screening each article by reading the abstract and title, 41 articles were selected based on the full-text eligibility assessment, title, and abstract, and if they were full-text articles. Thus, nine articles were removed as they met the exclusion criteria and were irrelevant to the research question. This resulted in 31 articles being selected as they met the inclusion criteria. After conducting a quality assessment of the full-text articles, four articles were removed as they were low in quality due to not having the relevant context and not answering the research questions. This resulted in 28 articles being used for this review. Figure 1 shows the PRISMA diagram, a graphical representation of the data extraction process.

QUALITY RANKING AND ASSESSMENT

According to Xiao and Watson (2019), when conducting an SLR, appraising quality is essential to ensure a quality SLR in the same line. Okoli (2015) states that an SLR should have a quality appraisal that ensures that all included articles are assessed for their quality to ensure confidence in the results obtained from the research. The author further states that all articles excluded due to insufficient quality should be assessed and not scored further. However, they should not be eliminated from the study. They should state why they were excluded. According to Xiao and Watson (2019), the results of the articles are scored according to how they appropriately meet the dimensions stated. These scores are then summarized to assess the quality of the SLR. The specific criterion for assessing the quality is according to Xiao and Watson's (2019) checklist for assessing the quality of qualitative studies. Table 2 shows the dimensions/criteria listed. The table lists the total number of articles with low, medium, and high-quality levels. Each article was given a score of (1) for low, (2) for medium, and (3) for high according to the dimensions mentioned. Articles that received a higher score were higher in quality, and articles that received a lower score were lower.

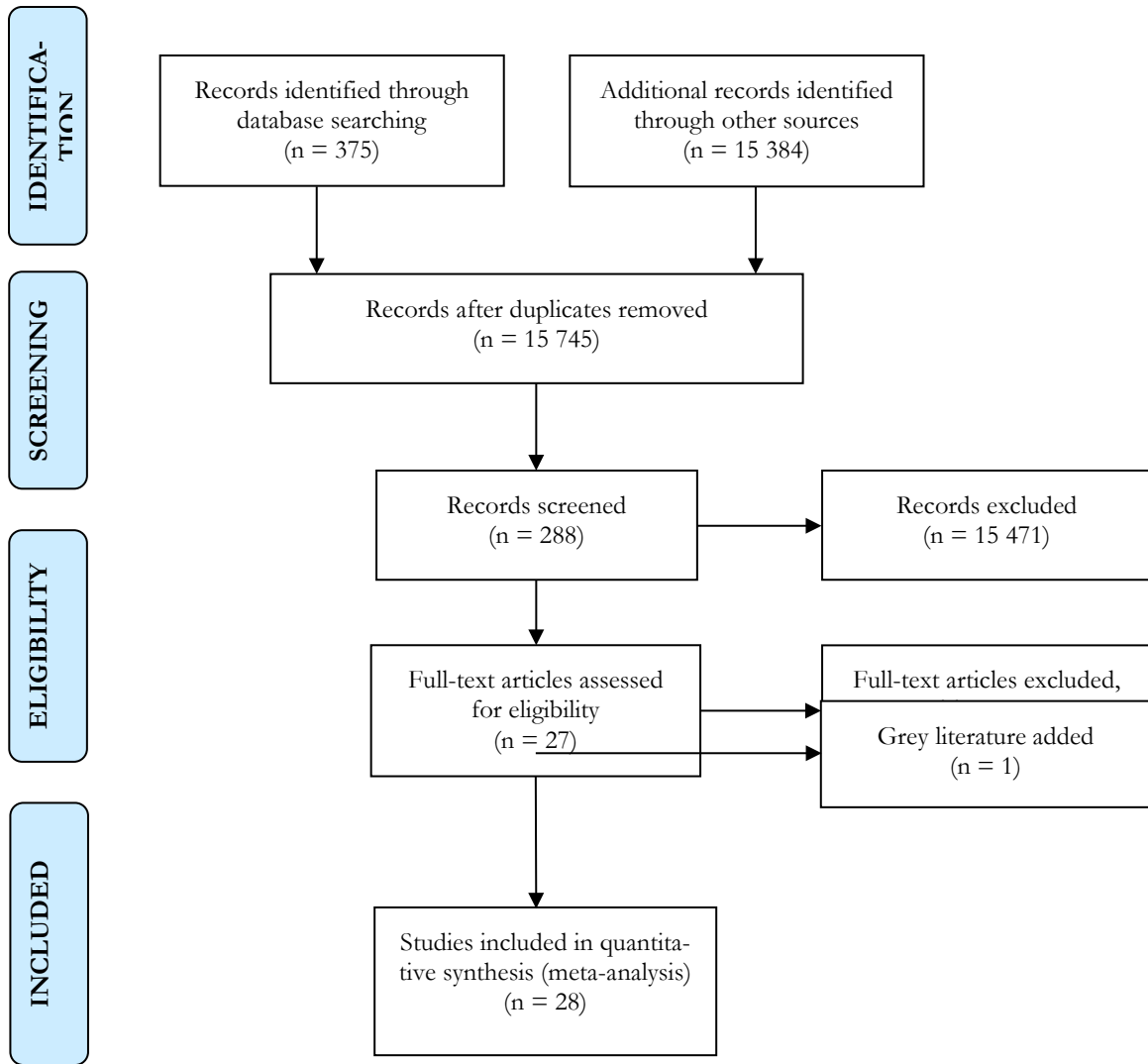


Figure 1. PRISMA Diagram

Table 2. Summary of quality assessment

Number	Quality Criteria/Dimensions	Low (1)	Medium (2)	High (3)
1.	How appropriate is the research design for addressing the question or sub-questions of this review?	3	6	23
2.	To what extent can the study findings be trusted in answering the study's question(s)?	-	2	30
3.	How specific is the context of the study represented?	8	7	17

The first dimension assessed was “How appropriate is the research design for addressing the question or sub-questions of this review?” Low (1) means that the paper was not relevant to the research question, Medium (2) means the paper somewhat relates to answering either the subquestions or main questions, and High (3) means the paper answers the research question. The reason for this

dimension is to assess whether the relevant articles would be able to answer the research questions stated and provide insights into understanding the research conducted. The next dimension is “To what extent can the study findings be trusted in answering the study’s question(s)?” Low (1) means that the paper cannot be trusted in giving accurate insights towards the research, Medium (2) means that the paper can be somewhat trusted in answering the research, High (3) means the paper can be trusted to answer the research question. This dimension was used to ensure that the data in the paper was accurate and could provide reliable results in answering the research question. The last dimension is “How specific is the study context represented?” Low (1) means that this paper is not relevant to the context of the study, or the context is not mentioned in the paper, Medium (2) means the context is described in that paper, and High (3) means the context is clearly stated, and the data is shown. This dimension was selected to ensure the context is within the context of the internet and social media. After the relevant eligibility of articles that scored a 1 in quality criteria 1 and 1 in quality criteria 3, they were excluded from the research as they did not answer the research question and were in an entirely different context, irrelevant to the main research question. Figure 2, in form of a bar graph illustrates the overall quality of the research articles used in this study.

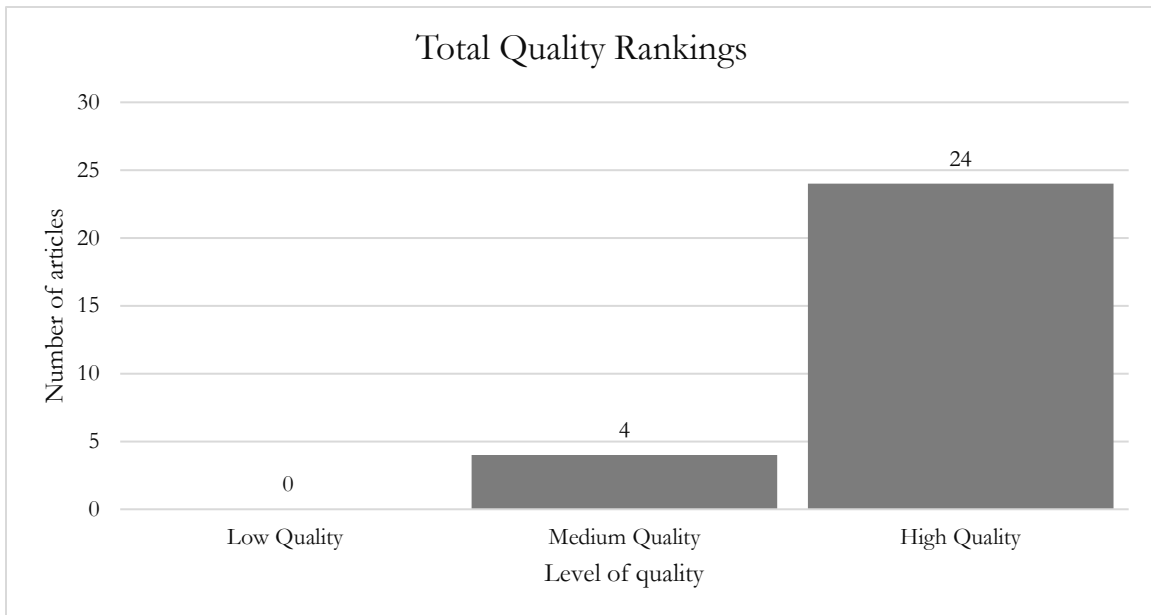


Figure 2. Total quality of articles

DATA ANALYSIS

A thematic analysis was used to answer the main research question. The field of deep fake technology has a variety of aspects, challenges, and issues related to it. A thematic analysis’s main benefits are its flexibility and its potential for a researcher to gain a better analysis in their field of research (Kiger & Varpio, 2020; Lester et al., 2020). According to Braun and Clarke (2024), conducting a thematic analysis starts with knowing the data, organizing it for analysis, and getting familiar with it. They suggest repeatedly reading the data to understand and writing ideas of what one recognizes in the data. Step two is to create codes by identifying and noting familiar patterns that are recognized in the data. Now, one needs to identify themes by finding “umbrella” terms that best describe the codes, reviewing those themes, and then producing an analysis (Braun & Clarke, 2024; Kiger & Varpio, 2020; Lester et al., 2020).

This study followed procedures recommended by Braun and Clarke (2024) for conducting thematic data analysis. Each article that was chosen and succeeded in the quality assessment was first read and understood. After that, words that relate to or describe the sub-research question were selected as the study's codes. All the codes identified were then written down on a spare sheet of paper. All codes that were repetitive or similar were eliminated. The codes that were left were then categorized into themes. Themes are words that best describe all the codes related to them. Then, these themes and their codes were reviewed and analyzed to ensure they were appropriate for the study. The codes were then recorded onto the matrix and combined to form a thematic map (Figure 3) that best illustrated all the relationships between the themes and the codes according to the primary and sub-research questions.

RESULTS

THEMATIC MAP OF FINDINGS

Each sub-research question prompted the codes to be identified within the research paper assessed. Using these codes, common themes were assigned for each research question, as illustrated in the thematic map below (Figure 3).

This thematic map assists in building the foundation for the discussion of the relevant findings. The first sub-research question explored the different types of fake news identified and categorized into two themes: categories of fake news and the forms of fake news. Most codes, such as misinformation and disinformation, were recurring in the first theme. The forms of fake news, such as sponsored content, propaganda, and clickbait, were typical examples of fake news. All the codes are about how information is formed into different types of fake news and how these codes fall under misinformation or disinformation, depending on the user's intentions. Sub-research question two was categorized into two themes: the advantages and disadvantages of deep fake technology. This theme explored all the benefits and drawbacks of deep fake technology, illustrating both the negative and positive effects of this technology.

In the third sub-research question, three themes were identified. Theme 1, human and civil rights, stems from the codes that elaborate on free speech as an ethical dilemma, as deep fakes allow one to express their creative ability and intellectual property, and codes such as invasion of privacy. The next theme is mental and physical health, expressing how deep fakes can decrease the cognitive ability of individuals because no individual can fully detect deep fake videos. Codes such as false memories and individual memories can be altered to believe true lies, and human deception was among the codes for this theme. The last theme identified is national security, which stems from codes that illustrate how deep fakes can cause political unrest between countries, cause significant events such as elections to be tampered with, and cause a decreased civil trust. The last sub-research question identified the theme of societal concerns. This theme is tied to codes that describe all the issues that arise when deep fakes are a new form of fake news in society. In addition, the codes that describe all the effects and impacts that deep fake technology can have on society.

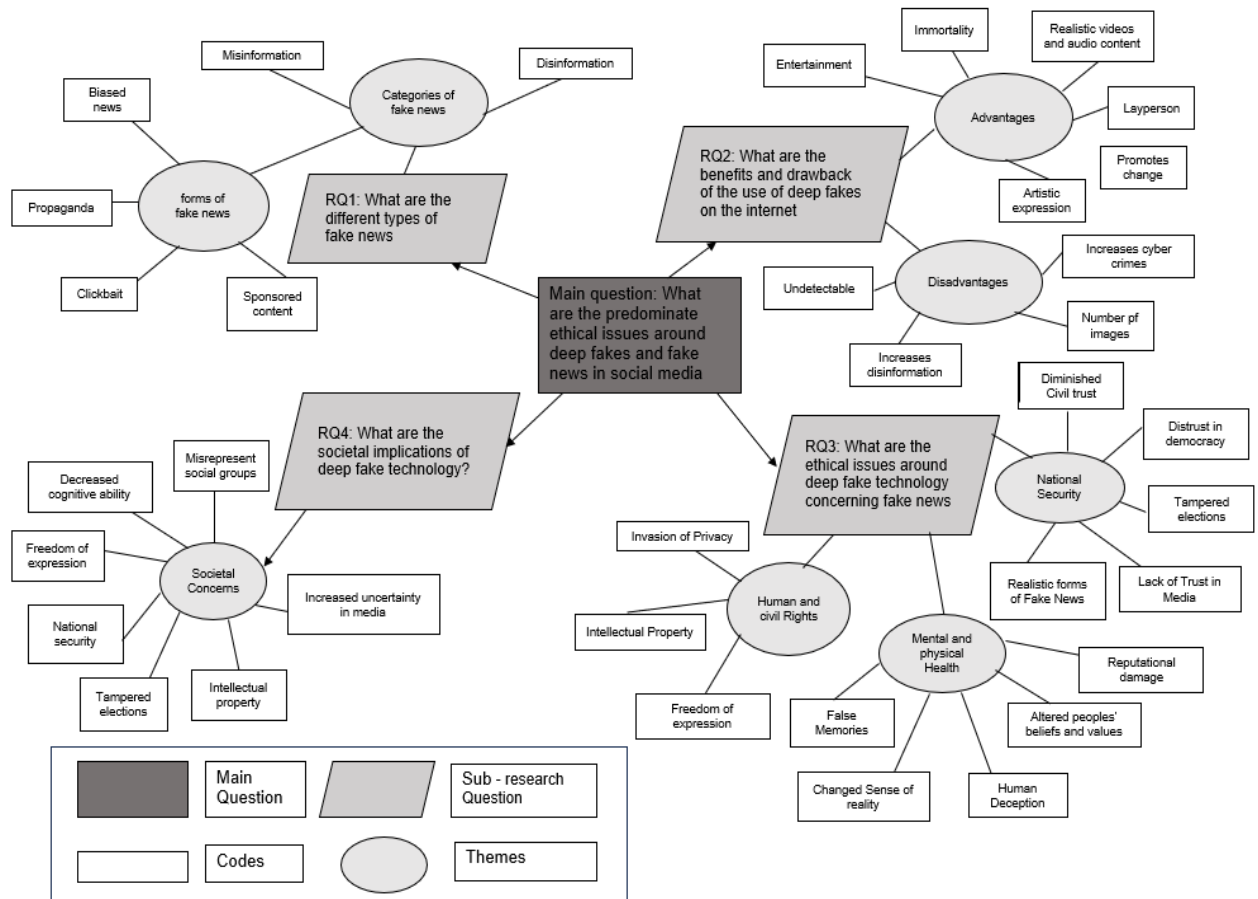


Figure 3. Map of thematic findings

DISCUSSION OF FINDINGS

The discussion of findings is done in alignment with the research question answered. This study seeks to answer the following research questions.

WHAT ARE THE TYPES OF FAKE NEWS?

This study further revealed that information can fall under two categories: misinformation and disinformation. As supported by Lazer et al. (2018), misinformation is misleading information about the world, but does not have the intention of causing harm. Previous scholars further argue that fake news is a subcategory of misinformation. It becomes complex to check for information credibility, as there are multiple sources of such information. Disinformation is information that is spread to mislead individuals, for example, deep fake videos (Helmus, 2022). Lazer et al. (2017) add that misinformation in today's era travels quickly due to social media platforms, as individuals share information across the globe. Interestingly, these fake news and false information spread faster than truthful information, and they hence state that the motivation for spreading false news, such as the social impact, can create and change the users' perceptions of a specific topic, thus altering their perception of a person based on false truths (Meel & Vishwakarma, 2020; Vaccari & Chadwick, 2020).

Fake news comes in various forms (Botha & Pieterse, 2020), such as clickbait, when news stories are fabricated to get more people to visit the website. This form of fake news is satire or parody, which is created to entertain people with no intention to harm. This type of news is different from propaganda, which is news that is created with the intent of misleading a particular audience. Propaganda is

another form of fake news that is fabricated and manipulated, through alteration of original news to deceive the readers, and has the potential to cause harm. In addition, sponsored content is a form of fake news edited to look like editorial content; however, it can be misleading to readers (Botha & Pieterse, 2020). Furthermore, biased news is a piece of news that relies on the reader's beliefs to mislead them. Lastly, sloppy journalism is news written using unreliable information or sources that cannot be trusted and potentially mislead readers. They further mention that these forms of fake news are usually used for financial gain or to promote specific ideas on topics. Misinformation and disinformation can therefore be amplified by new and existing technologies, such as deep fake technology (Botha & Pieterse, 2020), which can be detrimental in the world of information (Helmus, 2022; Lazer et al., 2017). The following section discusses the benefits and challenges of deep fakes in society.

HOW DOES THE USE OF DEEP FAKES BENEFIT AND CHALLENGE DIGITAL SOCIETY?

The previous section provided answers on the types of fake news. This section answers the research question on the benefits and challenges of deep fakes on society. The research question was answered by highlighting deep fake technology's positive and negative impact on society. Deep fake technology has had its fair share of scrutiny among researchers due to its adverse effects on individuals who became victims of deep fake technology (Chesney, 2022; de Ruiter, 2021; Meskys et al., 2020). However, the findings show that deep fakes have benefits to society. For example, Botha and Pieterse (2020) and Brooks (2021) argued that deep fake technology affords an advantage in that it is easy to use. Any person with no technological background can create a deep fake audio or video that is convincing and not easily detectable to machines or humans (Smith & Mansted, 2020). Deep fake technology can therefore be used for entertainment purposes, such as in sartorial or meme content, which is created as a form of social commentary (Botha & Pieterse, 2020; Farid, 2022). Another entertainment feature that deep fakes can create is upgrading video dubbing/synthesia for videos or movies. This was seen when the famous footballer David Beckham's video was recorded using nine different languages to fight against malaria, which had a positive social impact (Farid, 2022).

Deep fake technology can also be used to restore speech to individuals who have lost it due to some trauma (Farid, 2022). For example, if an actor loses his/her voice due to throat cancer, deep fakes can be used to convert the actor's original voice into sound from just typing text on a device (Farid, 2022). In addition, Farid (2022) argues that deep fakes can be used to create short animations of deceased individuals and digital avatars. Similarly, Gamage et al. (2022) reiterate that deep fake technology opens opportunities in the technical space of artificial intelligence. They further state that deep fakes can revolutionize the customer service industry and online course offerings, as this technology is versatile.

Despite the useful applications of deep fakes, the technology can be used to manipulate or alter videos from dash cams, nanny cams, body cams, and even car cams. It allows for tampering with any form of video, which can have an influential effect across multiple domains (Brooks, 2021). Another downside to deep fake technology is that there is a need for thousands of photos, also called training data, to create a deep fake; hence, celebrities and politicians are the most victims of deep fake content (Helmus, 2022). In courtrooms and police investigations, they rely on media content such as audio and video to provide evidence. Therefore, the prevalence of deep fake videos and audios can lead to tampered evidence that is not reliable in proving one's case. Videos must be thoroughly examined to detect tampering before being used in court, which is a serious downfall (Koopman et al., 2018).

This study demonstrates that deep fakes have the potential to cause what scholars termed the "liars dividend". The liars dividend implies that people have lost all efforts to detect propaganda or misinformation, and they lose trust in all sources of information, which can cause a detrimental effect on society. The liars dividend gives room for a diminished civil trust, which can lead to a societal collapse (Smith & Mansted, 2020). Furthermore, Smith and Mansted (2020) emphasized that deep fake technology can be used as a tool in cybercrimes, such as advanced phishing attacks, and fueling

online propaganda by mimicking people in power. The next section discusses findings on the predominant ethical issues around deep fake technology and fake news on social media.

WHAT ARE THE PREDOMINANT ETHICAL ISSUES AROUND DEEP FAKE TECHNOLOGY AND FAKE NEWS ON SOCIAL MEDIA?

Deep fake technology has raised ethical concerns that can pose significant risks to individuals and societies. The first ethical concern is the threat deep fakes pose to personal safety and privacy (Brooks, 2021; Chesney & Citron, 2019). Privacy is the individual's right from unwanted disclosure of personal information by any person, government, or corporation (Barrett-Maitland & Lynch, 2020). For example, the use of deep fake technology on famous actors such as Gal Gadot and Emma Watson's faces on pornographic actors was nonconsensual (Chesney & Citron, 2019). This act led to the violation of privacy, invasion of the actors' natural rights, and violation of sexual privacy (Barrett-Maitland & Lynch, 2020).

Furthermore, such unethical use of deep fake technology leads to a person's emotional abuse and reputational damage through cyberbullying, especially amongst teenagers, thereby threatening their career prospects, acceptance in society, and in some instances, safety, as supported by Brooks (2021). Social media platforms provide the environment for such deep-fake audios and videos to go viral. In the case of deep fake pornographic videos, it has the potential to marginalize communities and negatively impact women, as their sexual privacy is violated (Brooks, 2021; de Ruiter, 2021; Vaccari & Chadwick, 2020). Despite deep fakes having negative effects on the individual and society levels, this technology has the potential to sabotage business organizations by damaging their reputation. Damaged reputational and jeopardized corporate image have a negative impact on companies' business opportunities and profits (Chesney & Citron, 2019; Farid, 2022).

On a broader scale, deep fake technologies can impose a bigger negative impact on national security, especially during sensitive negotiations or conflicts between countries. For example, if a deep fake video were released in Gaza showing an Israeli soldier murdering a Palestinian child (Chesney & Citron, 2019), this would cause violent civil unrest, which can have devastating consequences for regional stability. Similarly, during critical times such as national elections, deep fakes can be used to sabotage results announcements and post-election stability and security, especially when deep fake announcements are released at the right time (Chesney & Citron, 2019; de Ruiter, 2021; Farid, 2022; Meskys et al., 2020). Deep fakes have the potential to erode trust in civil and democratic institutions. According to Chesney and Citron (2019), a good democracy requires truth and facts so that citizens can consider, debate, and ponder over topics for the good of the country, and a deep fake can jeopardize the possibility of having these conversations. However, the use of deep fakes can distort reality through disinformation that disrupts the politics of democracies, thus instilling false beliefs and harming social relations and trust (de Ruiter, 2021).

Another ethical issue is the formation of false memories, which presents ethical concerns that are up for discussion in this study. According to Liv and Greenbaum (2020), memories could be implanted into people's minds. These can influence humans' cognitive abilities as their memories can be altered for a particular motive. For example, during election time, people in power (e.g., politicians) can release deep fake videos to drive a particular narrative without providing specific details of what happened. This video is inaccurate and customized for that agenda. So, when the election occurs, the electorate will mostly remember and be influenced by the inaccurate video previously shared that disregards the other candidate (Liv & Greenbaum, 2020).

The ethical issues that have an impact on moral issues such as privacy, human and civil rights and national security were the main ethical concerns with deep fakes in the context of fake news (Brooks, 2021; Silbey & Hartzog, 2019), violate their freedom of speech and cause people to lose faith in their country's democracy. Deep fakes add to online disinformation, which is parallel to fake news, which is misinformation, and the use of deep fakes can create uncertainty among individuals and society,

and tend to have a decreased trust in their government (Vaccari & Chadwick, 2020). The following section discusses the social implications of using deep fake technology.

WHAT ARE THE SOCIETAL IMPLICATIONS OF USING DEEP FAKE TECHNOLOGY?

The last and final research question in this paper is to interrogate the societal implications that deep fake technology can have on society and assist policymakers in mitigating these implications. The ability of deep fake technology to create false realities in individuals' minds raises ethical concerns about the implantation of false memories (Liv & Greenbaum, 2020). Deep fake videos or audio recordings used or exploited to create a story or fabricate an event in one's favor can cause a generation to believe an event in an altered way, instead of the actual event (Liv & Greenbaum, 2020). Individuals normally use videos and pictures to prove or tell the truth about specific incidents, and we prove their validity through these forms of media. However, with the utilization of deep fake technology, it becomes difficult to verify the authenticity of video and audio recordings (de Ruiter, 2021). This can cause concerns in court cases, as pictures and videos have been primarily used to prove justice, leading to unjust persecution (Hancock & Bailenson, 2021).

Findings further revealed that deep fake technology can cause compromised national security, and war declarations or messages that are fabricated on social media, which have the potential to lead to political unrest (Brooks, 2021; de Ruiter, 2021; Hancock & Bailenson, 2021). Another ethical issue we face as a society in the context of deep fakes is that there is no accountability for the people who release deep fake videos (Hancock & Bailenson, 2021). This imperiled our society as people escape responsibility for the harmful, deep, and fake content. Therefore, we need governments to put tech people and lawyers together to combat the problem and mitigate the risks that deep fakes can have on our society (Chesney & Citron, 2019; Hancock & Bailenson, 2021).

CONCLUSION AND RECOMMENDATIONS

This paper discusses the various risks of deep fake technology and fake news perpetuated through social media on society. Some of the challenges revealed by the study pertain to individuals' exploitation and sabotage, blackmail, privacy violation, civil trust and unrest, and intellectual property rights. From the systematic literature review, it becomes evident that deep fakes can potentially ruin the reputation of individuals, companies, and nations on a larger scale. Deep fakes can cause political unrest as they can misrepresent certain marginalized groups and provide misleading information that can cause threats such as decreased civil trust in institutions and democracies, and even war. These ethical concerns underscore the pressing need for ethical considerations to ensure the safety and well-being of individuals and our society. Therefore, this paper recommends that governments should hold perpetrators accountable for releasing deep fake technologies, and there needs to be some form of responsibility through regulatory frameworks. In addition, technologists and lawyers must come together to propose guiding policies and strategies that mitigate the risks that deep fake technologies bring.

IMPLICATIONS FOR THEORY AND PRACTICE

The study has implications for theory and practice. From a theoretical standpoint, this study challenges traditional concepts of media literacy and individuals' comprehension of truth as deep fake technologies blur the lines between fake and real. From a practice standpoint, it is anticipated that technology developers, governments, content creators, and other consumers of deep fake technology will benefit from this study. Measures should be implemented to safeguard citizens from the harmful effects of deep fake technology through stronger regulatory frameworks, ethical guidelines, and corporate accountability.

FUTURE RESEARCH

Future research can focus on various aspects of deep fake technology, such as a comparative analysis of this technology in developing countries versus developed countries. Future research can also be conducted on this topic to understand the intricacies and ethical dimensions of deep fakes and artificial intelligence in the domains of law and politics. For example, most research has highlighted the limitations of the law on the protection of victims of deep fakes. Future research may expand the focus on the phenomenon of deep fakes to include the possible impact of deep fake technology on cybersecurity.

REFERENCES

-
- Albahar, M., & Almalki, J. (2019). Deepfakes: Threats and countermeasures systematic review. *Journal of Theoretical and Applied Information Technology*, 97(22), 3242-3250.
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-236. <https://doi.org/10.1257/jep.31.2.211>
- Barrett-Maitland, N., & Lynch, J. (2020). Social media, ethics and the privacy paradox. In C. Kalloniatis & C. Travieso-Gonzalez (Eds.), *Security and privacy from a legal, ethical, and technical perspective*. IntechOpen. <https://doi.org/10.5772/intechopen.90906>
- Botha, J., & Pieterse, H. (2020, October). Fake news and deepfakes: A dangerous threat for 21st century information security. *Proceedings of the 15th International Conference on Cyber Warfare and Security, Islamabad, Pakistan*. <https://doi.org/10.1109/iccws48432.2020.9292375>
- Braun, V., & Clarke, V. (2024). Thematic analysis. In F. Maggino (Eds.), *Encyclopedia of quality of life and well-being research* (pp. 7187-7193). Springer. https://doi.org/10.1007/978-3-031-17299-1_3470
- Brooks, C. F. (2021). Popular discourse around deepfakes and the interdisciplinary challenge of fake video distribution. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 159-163. <https://doi.org/10.1089/cyber.2020.0183>
- Chesney, R. (2022). *Disinformation on steroids: The threat of deep fakes*. Council on Foreign Relations. <https://www.jstor.org/stable/resrep29943>
- Chesney, R., & Citron, D. K. (2019). 21st century-style truth decay: Deep fakes and the challenge for privacy, free expression, and national security. *Maryland Law Review*, 78(4), Article 5. <https://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=3834&context=mlr>
- de Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology*, 34(4), 1311-1332. <https://doi.org/10.1007/s13347-021-00459-2>
- Dhiman, B. (2023). *Ethical issues and challenges in social media: A current scenario*. <https://doi.org/10.2139/ssrn.4406610>
- Drucker, A. M., Fleming, P., & Chan, A. W. (2016). Research techniques made simple: assessing risk of bias in systematic reviews. *Journal of Investigative Dermatology*, 136(11), e109-e114. <https://doi.org/10.1016/j.jid.2016.08.021>
- Farid, H. (2022). Creating, using, misusing, and detecting deep fakes. *Journal of Online Trust and Safety*, 1(4). <https://doi.org/10.54501/jots.v1i4.56>
- Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E., & Sasahara, K. (2022). Are deepfakes concerning? Analyzing conversations of deepfakes on Reddit and exploring societal implications. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1-19). Association for Computing Machinery. <https://doi.org/10.1145/3491102.3517446>
- Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 149-152. <https://doi.org/10.1089/cyber.2021.29208.jth>
- Harris, D. (2018). Deepfakes: False pornography is here and the law cannot protect you. *Duke Law & Technology Review*, 17(1), 99-127.

- Helmus, T. C. (2022). *Artificial intelligence, deepfakes, and disinformation: A Primer*. RAND Corporation. <https://www.jstor.org/stable/resrep42027>
- Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher*, 42(8), 846-854. <https://doi.org/10.1080/0142159X.2020.1755030>
- Koopman, M., Rodriguez, A. M., & Geradts, Z. (2018, August). Detection of deepfake video manipulation. *Proceedings of the 20th Irish Machine Vision and Image Processing Conference, Belfast, Northern Ireland*, 133-136.
- Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? assessment and detection. PsyArXiv. <https://doi.org/10.48550/arXiv.1812.08685>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096. <https://doi.org/10.1126/science.aao2998>
- Lazer, D. M. J., Baum, M. A., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., & Mattsson, C. (2017). *Combating fake news: An agenda for research and action*. Shorenstein Center on Media, Politics and Public Policy. <https://apo.org.au/node/76233>
- Lester, J. N., Cho, Y., & Lochmiller, C. R. (2020). Learning to do qualitative data analysis: A starting point. *Human Resource Development Review*, 19(1), 94-106. <https://doi.org/10.1177/1534484320903890>
- Liang, B., Wang, Z., Huang, B., Zou, Q., Wang, Q., & Liang, J. (2023). Depth map guided triplet network for deepfake face detection. *Neural Networks*, 159, 34-42. <https://doi.org/10.1016/j.neunet.2022.11.031>
- Liv, N., & Greenbaum, D. (2020). Deep fakes and memory malleability: False memories in the service of fake news. *AJOB Neuroscience*, 11(2), 96-104. <https://doi.org/10.1080/21507740.2020.1740351>
- Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, 153, 112986. <https://doi.org/10.1016/j.eswa.2019.112986>
- Meskys, E., Kalpokiene, J., Jurcys, P., & Liaudanskas, A. (2020). Regulating deep fakes: Legal and ethical considerations. *Journal of Intellectual Property Law & Practice*, 15(1), 24-31. <https://doi.org/10.1093/jiplp/jpz167>
- Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37. <https://doi.org/10.17705/1cais.03743>
- Pérez Dasilva, J., Meso Ayerdi, K., & Mendiguren Galdospin, T. (2021). Deepfakes on Twitter: Which actors control their spread? *Media and Communication*, 9(1), 301-312. <https://doi.org/10.17645/mac.v9i1.3433>
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, 8(3), 171-188. <https://doi.org/10.1089/big.2020.0062>
- Silbey, J., & Hartzog, W. (2019). The upside of deep fakes. *Maryland Law Review*, 78(4), Article 8.
- Smith, H., & Mansted, K. (2020). *Weaponised deep fakes: National security and democracy* (pp. 11-14). Australian Strategic Policy Institute. <https://www.jstor.org/stable/resrep25129.7>
- Statista. (2024). *Share of internet users who use social networks worldwide from 2019 to 2024*. <https://www-statista.com/statistics/1202737/social-network-users-worldwide/>
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1). <https://doi.org/10.1177/2056305120903408>
- Wagner, T. L., & Blewer, A. (2019). "The word real is no longer real": Deepfakes, gender, and the challenges of ai-altered video. *Open Information Science*, 3(1), 32-46. <https://doi.org/10.1515/opis-2019-0003>

- Wahl-Jorgensen, K., & Carlson, M. (2021). Conjecturing fearful futures: journalistic discourses on deep-fakes. *Journalism Practice*, 15(6), 803-820. <https://doi.org/10.1080/17512786.2021.1908838>
- Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1), 93-112. <https://doi.org/10.1177/0739456X17723971>
- Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019, July). Unsupervised fake news detection on social media: A generative approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1), 5644-5651. <https://doi.org/10.1609/aaai.v33i01.33015644>
- Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5), Article 109. <https://doi.org/10.1145/3395046>
- Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019). Fake news: Fundamental theories, detection strategies and challenges. *Proceedings of the 12th ACM International Conference on Web Search and Data Mining* (pp. 836-837). Association for Computing Machinery. <https://doi.org/10.1145/3289600.3291382>

AUTHORS



Singarila Somanje is a graduate with a strong foundation in Information Systems, holding a BCom Honours in Information Systems from the University of the Witwatersrand (WITS) and a BCom in Information Systems from the University of Johannesburg (UJ). With a solid academic background, she has further honed her IT ethics and compliance skills by developing a deeper understanding of IT audits, risk assessments, and compliance reviews through hands-on experience as a Junior Analyst at KPMG. In this role, she worked closely with IT audit teams, gaining practical insights into industry practices and compliance standards of IT.



John Mangundu has a Doctor of Philosophy (PhD) in Information Systems and Technology. His defined areas of specialization are information technology (IT) governance, cybersecurity, and information and communication technologies (ICTs) in education. Mangundu has published various peer-reviewed research papers and a book in his areas of specialization. In addition, John Mangundu has presented various research papers at national, regional, and international research conferences. Mangundu has thirteen years of lecturing experience gained between Zimbabwe and South Africa. Dr John Mangundu is a Senior Lecturer at the University of the Witwatersrand, South Africa.