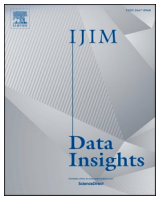
International Journal of Information Management Data Insights 4 (2024) 100286

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei



Extending the theory of information poverty to deepfake technology

Walter Matli

University of South Africa, South Africa

ARTICLE INFO

Keywords:
Deepfake technology
Information poverty theory
Artificial intelligence (AI)
Synthetic media
Societal implications
Technological advancements

ABSTRACT

The advent of deepfake technology has introduced complex challenges to the information technology landscape, simultaneously presenting benefits and novel risks and ethical considerations. This paper delves into the evolution of deepfakes through the prism of information poverty theory, scrutinising how deepfakes may contribute to a growing information access/use inequality. The research focuses on the risks of misinformation and the ensuing expansion of digital divides, particularly when manipulative media could delude individuals lacking access to legitimate information sources. The study outlines the potential exacerbation of information asymmetries and examines the societal implications across various demographics. By integrating an analytical discussion on the risks associated with deepfakes, the study aligns the observed trends with the theoretical underpinnings of information poverty. As part of its contribution, the paper offers actionable policy-making recommendations and educational strategies to combat the proliferation of harmful deepfake content. The article aims to ensure a more equitable distribution of authentic information and foster media literacy. Through a multifaceted approach, this study endeavours to provide a foundational understanding for stakeholders to navigate the ethical minefield posed by deepfakes and to instil a framework for information equity in the digital era. The article provides critical insights into the discourse on deepfake technology and its relation to information poverty, underscoring the urgent need for equitable access to informed digital spaces. As deepfake technology evolves and more data emerges, a societal demand exists for comprehensive knowledge about deepfakes to promote discernment, decision-making and awareness. Policymakers are tasked with recognising the significance of widening access to sophisticated information technologies whilst addressing their negative repercussions. Their efforts will be particularly crucial for disseminating knowledge about deepfakes to those with limited or non-existent information and communication awareness and infrastructures. Learning from past successes and failures becomes pivotal in shaping effective strategies to address the challenges posed by deepfakes and fostering accessible, informed digital communities.

1. Introduction

Deepfake technology leverages deep learning algorithms to create highly realistic and manipulated videos or images that depict events or individuals in ways that never actually occurred. Furthermore, text, data, numbers, and messages can be altered to reflect specific viewpoints or arguments. The rapid development and deployment of deepfake technology poses significant challenges and opportunities in the digital age. The proliferation of deepfake technology presents a significant challenge, affecting society, personal privacy, and many other aspects. The potential harm of deepfakes to democracy, public discourse, and societal stability necessitates a comprehensive and proactive approach. As deepfake technology advances, staying vigilant and proactive is imperative in countering its harmful effects on individuals and society.

The digital age has ushered in an era of unprecedented technological advancement, but it has also unveiled a andora's box of new threats. Among the most insidious is the rise of deepfakes – sophisticated AI-powered forgeries capable of bending reality (Whyte, 2020). By leveraging deep learning algorithms, deepfakes can manipulate videos, images, text, and even data, to depict events or individuals in ways that never occurred, often with alarming realism (Lyu, 2020). This ability to manipulate information mirrors the capabilities seen in other domains where deep learning has revolutionised analysis, such as in medical imaging, where convolutional neural networks can now identify different types of tumours with remarkable accuracy (Mijwil et al., 2023).

This ability to fabricate reality easily and precisely presents a multifaceted challenge to society (Pawelec, 2022). The proliferation of deepfakes threatens personal privacy, democratic processes, and the very foundation of trust in information (Nguyen et al., 2022). As

E-mail address: matliw@unisa.ac.za.

https://doi.org/10.1016/j.jjimei.2024.100286

Available online 24 September 2024

deepfake technology becomes increasingly sophisticated and accessible, a comprehensive and proactive approach is paramount to mitigate its potential harms (Buo, 2020).

Borji (2022) focuses on evaluating Generative Adversarial Networks' (GANs) performance, providing insights into their effectiveness and suggesting future research directions. Deepfakes, created using advanced machine learning techniques such as GANs, have the potential to exacerbate information inequality, contributing to what is known as information poverty. Advances in deep learning models have made the creation of deepfakes more accessible, raising concerns about potential misuse (Chen et al., 2021). The rapid evolution of deepfake technology has made it increasingly challenging to differentiate between authentic and fabricated content (Firc et al., 2023). Consequently, the potential for deepfakes to erode public trust, damage individual reputations, and influence public opinion has become a pressing concern (Seow et al., 2022).

A systematic literature review approach has been adopted in this paper, drawing in particular on research in both the deepfake technology and information poverty fields, and using the suggestions provided by Varsha et al. (2024) in integrating the literature from the two source fields.

Existing literature highlights various gaps and challenges associated with deepfake technology and its detection. For instance, Harbinja et al. (2023) address the legal and ethical frameworks governing digital legacies, including the regulation of post-mortem data and digital personas. Liz-Lopez et al. (2024) and Qiu et al. (2023) underscore the complexity of detecting manipulated images and forgeries, noting that traditional methods often require large datasets and face significant obstacles in distinguishing between authentic and manipulated content. Guo et al. (2021) emphasise the need for robust detection mechanisms under complex scenarios involving post-processing operations.

In the context of multimedia data, Liz-Lopez et al. (2024) and Bonomi et al. (2021) highlight the evolution of forensic techniques, particularly the shift from single-frame detection methods to those exploiting temporal information in video sequences. Anwar et al. (2023) address the challenges of detecting complex forgery types like copy-move and splicing. The study by Lee et al. (2021) on digital image forensics highlights the necessity of distinguishing between authentic images and sophisticated forgeries, while Hamed et al. (2023) address the challenges in detecting fake news, pointing to the lack of comprehensive datasets and effective feature representation. Habbal et al. (2024) underscore the importance of balancing trust, risk, and security in artificial intelligence (AI) systems through transparent and ethical frameworks.

Addressing these challenges requires a multifaceted approach, as different studies cited below have agreed. Such an approach includes advanced detection technology, legislation and regulations, education and awareness programmes, and international collaboration (Dasilva et al., 2021). The multifaceted nature of this approach, involving various stakeholders, is crucial in developing effective strategies to detect and combat deepfakes (Godulla et al., 2021). The threat of deepfakes extends beyond individual deception and manipulation; it risks the integrity of democratic processes, the credibility of news and information, and society's overall stability (Reviglio, 2022). In the context of information poverty, the distorted information landscape created by deepfakes adds another layer of complexity, emphasising the need for a unified response. Through these diverse examinations, this study seeks to provide a comprehensive understanding of deepfake technology, highlighting its potential to exacerbate information inequality and guiding future research efforts in addressing these critical gaps. Therefore, this research is an initial effort to compensate for the deficiencies in the existing literature.

1.1. Problem definition

The proliferation of deepfake technology has heightened the risk of malicious abuse, making developing effective deepfake detection methods increasingly crucial (Hosler et al., 2021). This section aims to articulate the problem at hand, integrating the concept of information poverty to emphasise the broader societal implications of deepfakes. So, the rising concerns over deepfakes have prompted researchers to focus on developing methods for effective deepfake detection (Seow et al., 2022). In the context of information poverty, where the accessibility and reliability of information are critical, the potential harm posed by deepfakes adds further complexity to the existing challenges.

Developing robust methods for deepfake detection is not just a technical necessity but a crucial component in mitigating the risks and potential harms associated with deepfakes. The overarching risks encompass detrimental influences on public trust, manipulation of public opinion, and the erosion of the reliability of digital content. In a society already grappling with information poverty, where reliable information is scarce or not readily accessible, the consequences of deepfakes on trustworthiness, personal reputation, and public consensus become particularly significant. The indispensability of deepfake detection becomes evident in its role in combating the adverse societal effects of deepfakes. These effects extend beyond individual harm to broader consequences such as eroding public confidence, manipulating public sentiment, and undermining the reliability of digital content. In the realm of information poverty, where the veracity and accessibility of reliable information is already under threat, deepfakes exacerbate the challenges by introducing fabricated content that is challenging to distinguish from reality.

Efficiently negating the detrimental influences of deepfakes and safeguarding societal values from potential threats requires the diligent advancement of robust deepfake detection methods (Jeong et al., 2022). In a landscape where information poverty hinders the accessibility of reliable information, the need for effective countermeasures against deepfakes becomes paramount. The problem is multifaceted, involving the technical challenge of developing deepfake detection methods and the broader societal consequences of deepfake proliferation. In the context of information poverty, where reliable information is scarce or limited in terms of easy access, addressing the problem requires a comprehensive approach beyond technical advancements. It necessitates a concerted effort to protect individuals and society from the harmful impacts of deepfakes on trust, public opinion, and the reliability of digital content.

2. Related works

2.1. Deepfake technology overview

The increasing sophistication of deepfake technology is making it increasingly difficult to distinguish between real and manipulated content (Ahmed, 2021). This technology presents both significant challenges and opportunities, requiring a comprehensive and multifaceted response to mitigate its potential harms while exploring its beneficial applications.

Deepfakes employ advanced techniques such as Generative Adversarial Networks (GANs) to create realistic forgeries that can deceive even the most discerning viewers. Fu et al. (2023) delve into the complexities of face forgery detection, highlighting the diverse manipulation methods and varied forgery patterns as major challenges. Similarly, Guo et al. (2021) emphasise the difficulties in detecting fake face images, particularly under complex scenarios involving post-processing operations like compression, blurring, and scaling. Anwar et al. (2023) further identify challenges in existing image forgery detection methods, which struggle with complex forgery types like copy-move and splicing.

Despite advancements in face image forgery detection, there remains a scarcity of effective detectors for images subjected to challenging conditions (Guo et al., 2021). Addressing the deepfake threat requires integrating technological advancements, policy measures, media literacy, and collaboration among various stakeholders to minimise the

negative effects and safeguard the integrity of our digital landscape.

Promoting media literacy is crucial in educating the public about deepfakes, their potential harm, and how to verify the authenticity of digital content. Empowering individuals to critically analyse and question the information they encounter can significantly reduce the impact of deepfakes on society. This necessitates continuous investment in the research and development of advanced detection methods that can keep up with the evolving nature of deepfakes.

The rise of deepfake technology presents both challenges and opportunities (Yang et al., 2022). Proactively addressing the threats while exploring potential beneficial applications is crucial for society. A multifaceted approach should focus on developing detection algorithms, implementing stricter regulations, raising public awareness, and fostering stakeholder collaboration (Shi et al., 2023). Deepfake technology has the potential to deceive individuals and undermine public trust, even in scientific research (Shi et al., 2023).

A comprehensive approach to combating deepfakes combines technological advancements, regulatory measures, education and awareness programmes, and international cooperation. Public awareness and education are essential in combating deepfake threats, particularly in the context of information poverty. Promoting media literacy and teaching individuals to discern between real and manipulated content can empower them to make informed decisions and mitigate the spread of deepfakes. Industry leaders and companies must establish clear guidelines and ethical standards for creating and disseminating synthetic media, contributing to a more trustworthy information environment (Habbal et al., 2024; Harbinja et al., 2023).

Deepfake technologies have become a growing concern, as algorithms create increasingly realistic face images and videos (Liz et al., 2024; Shi et al., 2023). Studies such as Liu et al. (2023) and Shaheed et al. (2024) explore the difficulties in detecting GAN-generated images and presentation attacks on biometric systems, respectively, emphasising the need for more efficient and interpretable algorithms. These deepfakes can be used for malicious purposes, such as spreading fake news, compromising privacy, and damaging reputations. Detection of deepfakes is crucial to combating their harmful effects on individuals and society.

A comprehensive approach that includes deepfake detection technology, legislative measures, and regulations is essential in addressing the growing concern about deepfakes (Gong, 2020). Whittaker et al. (2023) point out the lack of guidance for managers in responding to synthetic media incidents in business settings.

To combat the threat of deepfakes effectively, a collaborative effort encompassing researchers, policymakers, technology companies, and the general public is necessary. This approach can develop robust detection tools, promote media literacy, and establish regulatory frameworks to curb the malicious use of deepfake technology (Reviglio, 2022). The potential consequences of deepfakes are far-reaching, impacting trust in institutions, damaging reputations, and manipulating public opinion (Kanwal et al., 2022). Governments should allocate resources to fund research and development of deepfake detection technology and establish legislation to discourage malicious use.

A multidisciplinary approach, leveraging the expertise and resources of key stakeholders, is essential for developing effective strategies to detect and combat deepfakes (Westerlund, 2019). Collaboration between technology companies and social media platforms is crucial to implementing effective solutions for detecting and preventing the spread of deepfakes. Social media platforms and technology companies must take responsibility for combating deepfakes by implementing robust content moderation policies and investing in AI systems that can detect and flag potentially harmful deepfake content. A multifaceted approach combining technological advancements, policy measures, education, and collaboration is necessary to effectively combat the spread of deepfakes and mitigate their potential harm to individuals and society.

AI technology, including deep learning algorithms, has the potential to revolutionise various industries and solve complex problems (G. Wang et al., 2022). However, it is crucial to recognise that AI can be misused and pose significant threats if not properly regulated. One effective way to address the issue of deepfakes is through the development and implementation of advanced detection technology. This technology could involve machine learning algorithms, such as convolutional neural networks, to analyse the spatiotemporal features in videos and identify any anomalies or inconsistencies that may indicate the presence of a deepfake.

While deep learning models draw inspiration from the brain, their artificial neurons operate quite differently from their biological counterparts. Biological neurons communicate through discrete electrical spikes, encoding information in both the timing and frequency of these spikes. This has led to the development of spiking neural networks (Tavanaei et al., 2019; Zheng et al., 2022), which aim to more realistically mimic the brain's computational mechanisms. Spiking neural networks (SNNs) not only offer a closer approximation of biological neural activity but also hold the potential for significantly greater energy efficiency due to their event-driven, asynchronous nature (Tavanaei et al., 2019).

Despite their potential, SNNs haven't yet reached the accuracy levels of traditional deep learning methods on all tasks, though the gap is closing (Yamazaki et al., 2022). SNNs even demonstrate superior performance in specific areas, particularly those involving spatio-temporal data. This trajectory of SNN development has significant implications for the issues of deepfakes and information poverty. On the one hand, more biologically realistic and efficient SNNs could lead to even more convincing deepfakes, further complicating the information landscape (Tavanaei et al., 2019). Conversely, the unique computational properties of SNNs might on the other hand hold the key to developing more robust deepfake detection mechanisms, ultimately helping to mitigate the spread of misinformation.

2.2. *Deepfake impacts on society*

The ramifications of deepfake technology are wide-ranging and can significantly impact various aspects of society. From political discourse and trust in institutions to personal privacy and online security, the rise of deepfake technology poses serious challenges. At present, there are several persistent challenges in identifying forgery detection. For instance, Fu et al. (2023) further explore the intricacies of face forgery detection, identifying the diverse manipulation methods and found forgery patterns as major challenges. In the same vein, other studies like Guo et al. (2021) explored the robust detection of fake face images, especially under complex scenarios that include post-processing operations such as compression, blurring, scaling, etc. In their study, Fu et al. (2023) found that traditional detection models face significant learning obstacles due to the varied forgery patterns produced by different manipulation techniques, such as deepfakes and neural textures. To address these challenges, a growing body of research focused on developing robust detection models capable of identifying deepfakes is essential (Khormali & Yuan, 2022). Additionally, efforts are now being made to establish legal frameworks and regulations to deter the malicious use of deepfakes. To counter the negative consequences of deepfakes, it is essential to invest in advanced deepfake detection technologies and tools. These detection models should be able to analyse various aspects of the content, such as facial movements, voice patterns, and inconsistencies in lighting and shadows. Fact checks and message validation techniques can also be deployed.

These challenges make it hard to develop robust systems that can reliably detect fake news and content across different contexts and media (Hamed et al., 2023). By harnessing the power of AI and machine learning, these solutions can continuously adapt and improve to keep up with evolving deepfake techniques. As Hamed et al. (2023) point out, the challenges in detecting fake news and content include issues like the lack of comprehensive and diverse datasets, difficulties in representing

features for machine learning models effectively, and complications in fusing data from various sources for more accurate detection. By deploying advanced detection algorithms and leveraging the collective efforts of researchers, policymakers, and technology companies, we can stay one step ahead of deepfake creators and minimise the potential harm caused by deepfake technology. In summary, the emergence of deepfake technology presents a significant threat in cyberspace (Firc et al., 2023). It is imperative that we take proactive measures to address and mitigate the negative impacts of deepfakes. By raising awareness, investing in research and development, implementing robust detection algorithms, and fostering collaboration between various stakeholders, we can effectively tackle the challenges posed by deepfakes and safeguard the integrity and trustworthiness of digital content. While deepfake technology can be entertaining and amusing in some instances, its potential for misuse is concerning. It is crucial to prioritise developing and implementing effective legal measures to prevent the malicious use of deepfakes.

In essence, the rise of deepfake technology presents serious challenges ranging from threats to privacy and online security to the spread of misinformation and impersonation (Chen et al., 2021). It is essential to invest in research and development of robust detection models, to establish legal frameworks and to deter malicious use, and promote collaboration between technology companies and social media platforms to effectively combat the harmful effects of deepfakes (Khormali & Yuan, 2022). The emergence of deepfake technology poses significant risks to our society, including the spread of misinformation and the potential for identity theft or political manipulation (Lin et al., 2022). Therefore, prioritising the development and implementation of detection and prevention methods is needed to safeguard against the misuse of deepfakes. Additionally, educating the public about deepfakes and their potential dangers is crucial in building awareness and resilience against digital manipulation.

The collaboration between Microsoft Corporation, the Partnership on AI coalition, and the research community through initiatives like the Deepfake Detection Challenge plays a vital role in developing and advancing detection methods for deepfakes. This collaboration allows for collective expertise, resources, and knowledge to be leveraged in the fight against deepfakes. By leveraging the power of technology, research, and collaboration, we can effectively combat the pervasive threat of deepfakes and protect the integrity of digital content. Deepfakes present a multifaceted threat to society, with implications for privacy, security, and public trust (Firc et al., 2023). Addressing these challenges requires a comprehensive approach that includes technological advancements in deepfake detection, the implementation of strong legal measures to deter malicious use, and public education and awareness campaigns to foster digital literacy and critical thinking skills (Khormali & Yuan, 2022). Furthermore, it is important for policymakers to prioritise the development and implementation of effective legal measures to prevent the malicious use of deepfakes. Therefore, there appears to be a consensus in the literature about the threats that deepfake technology poses and that a collaborative and comprehensive approach is required to combat such threats.

3. Theoretical perspective: information poverty theory and deepfakes

This study now explores the intersection of information poverty theory and the phenomenon of deepfakes, aiming to contribute valuable insights into the challenges posed by synthetic media within the broader context of information access, dissemination, and the potential harmful effects of deepfakes.

The research aims to make a significant contribution to the existing body of knowledge surrounding information poverty theory. The fundamental tenets of the theory remain unchanged, providing a conceptual framework to understand the multifaceted challenges related to the deprivation of the right to information. The study thus employs information poverty theory as a lens to delve into the intricate landscape of deepfakes, acknowledging that the theory's enduring relevance offers a solid foundation for understanding information disparities.

Information poverty is defined as the deprivation of the right to information, stemming from various factors such as lack of access, information overload, and self-imposed information deprivation (Sweetland, 1993). Britz (2004) adds depth to this definition, characterising it as a situation where individuals lack the necessary skills, abilities, or material resources to access, interpret, and apply information efficiently. Lingel and Boyd (2013), point out that the information poverty theory is often employed to elucidate challenges in obtaining information and community norms regarding information sharing or withholding.

According to Britz (2004), information poverty manifests as a deficiency in accessing crucial information, even when available, with individuals needing help to derive appropriate meaning. Goulding (2001:109) further describes it as an outcome of a divide between those with easy access to abundant information and those lacking the knowledge or resources to find it. The article explores the complex dynamics of misinformation, limited data for decision-making, and information poverty within the context of deepfakes.

The concern over inequality in information access emerged in the 1960s (Yu, 2006), revealing that certain individuals were disadvantaged due to a lack of essential skills and knowledge. Information inequality was influenced by factors such as inadequate literacy, digital skills, limited educational opportunities, and insufficient training. The transition to the information era, marked by advances in information and communication technologies (ICTs), exacerbated imbalances, perpetuating information inequality in society.

With the transition to the knowledge economy facilitated by ICTs (Freeman and Louçã, 2001), information poverty has taken on new dimensions. While ICTs have been transformative, at the same time deepening the gap between wealthy and impoverished nations, it is essential to recognise that information poverty extends beyond a single technology or digital divide. Cultural and linguistic diversity, educational levels, and the ability to access and benefit from information contribute to the multi-faceted nature of information poverty (Britz, 2004).

Today, information systems have proliferated but are still essentially designed to collect, process, store and distribute information via the essential components of people, tasks, structures and technology. What has recently ensued, however, is akin to an arms race between fake information creators and fake detection methods with the latter providing mitigation capabilities but also limitations.

3.1. Linking information poverty to deepfake technology

The continuous improvement and advancement of deepfake detection methods are essential to staying ahead of the rapidly evolving deepfake technology. This study underscores the importance of understanding the potential harm caused by deepfakes and the necessity of raising awareness among the general public about their existence and dangers (Jeong et al., 2022). By linking information poverty theory to the challenges posed by deepfakes, the research aims to provide a comprehensive understanding of the intricate dynamics between information disparities and the evolving landscape of synthetic media.

The concept of information poverty and the phenomenon of deepfakes are intertwined, as both relate to the challenges and distortions in the information landscape, albeit in different ways. Information poverty generally refers to the lack of access to reliable and accurate information, often as a result of various socio-economic and cultural factors. Deepfakes, on the other hand, involve the creation of highly convincing fake content, such as videos or audio recordings, as well as text and data, using advanced AI techniques.

Table 1 below illustrates how information poverty and deepfakes are related.

The combined impact of information poverty and deepfakes poses a

Table 1
Linking aspects of information poverty to deepfakes.

	Information Poverty Perspective	Deepfake Connection
Manipulation of information credibility	Individuals experiencing information poverty already face challenges in accessing credible information. The scarcity of reliable sources makes it difficult for them to discern accurate information from misinformation.	Deepfakes exacerbate this issue by introducing fabricated content that is challenging to distinguish from real information. The lack of media literacy and limited access to trustworthy sources in information-poverty scenarios makes individuals more susceptible to believing and spreading deepfake content.
Erosion of trust	Trust in information sources is crucial for individuals facing information poverty. When access to reliable information is limited, individuals may rely on a few sources, and any erosion of trust in these sources can have significant consequences.	Deepfakes have the potential to erode trust further by creating fake content that mimics real events or individuals. This undermines trust in digital media and the information ecosystem as a whole, especially for those already experiencing information poverty.
Impact on public opinion	In information-poverty scenarios, public opinion may be shaped by a limited set of information sources, potentially leading to a narrow or biased understanding of events.	Deepfakes can manipulate public opinion by creating false narratives or presenting individuals saying or doing things they never did. In environments where information is already scarce, the influence of deepfakes on public perceptions can be particularly potent.
Challenges in media literacy	Limited access to education and information resources can result in lower levels of media literacy, making individuals more vulnerable to misinformation. Reliance on social media platforms and mobile phones can allow deepfakes to spread rapidly without verification.	Detecting deepfakes requires a certain level of media literacy and technological awareness. In information-poverty contexts, individuals may need more skills and resources to critically evaluate the authenticity of digital content, increasing susceptibility to deepfake manipulation.
Societal stability and integrity	Information poverty can contribute to societal instability by limiting access to accurate information necessary for informed decision-making.	The spread of deepfakes threatens societal stability by introducing fabricated content that can influence political processes, public discourse, and social interactions.

significant challenge to maintaining the integrity of societal structures. Information poverty sets the stage for increased vulnerability to the harmful effects of deepfakes. The interplay between limited access to reliable information and the deceptive nature of deepfakes underscores the need for comprehensive strategies that address both socio-economic disparities in information access and the technological challenges posed by synthetic media.

4. The current study

In Gamage et al.'s (2023) view, deepfakes' impact on society is still being explored through research. Deepfakes have the potential to significantly impact society in both positive and negative ways. Much of the research and analyses focus on detection and prevention (Miranda-García et al., 2024; Byeon et al., 2024; Yadav & Vishwakarma, 2023; Liang et al., 2023). There may be this plethora of studies focusing on detecting and preventing deepfakes because when it comes to deepfake technology, it's important for society to stay informed about the latest research and developments in detection and prevention methods. Accordingly, researchers have made significant advancements in deepfake detection and prevention techniques. Similarly, other studies have focused on the technological advancements of deepfakes (such as Guo et al., 2023; Yungui et al., 2023; Raj et al., 2023). Technological advancements have played a significant role in the evolution of deepfake technology. These advancements have created more realistic and sophisticated deepfakes, making it increasingly difficult to distinguish them from genuine media. Much of the research has been reported from a developed-world perspective.

The available literature indicates that more attention needs to be devoted to the impact of deepfakes on people's lived experiences and situations from a developing country perspective. In this context, the general purpose of the paper is to explore and analyse the development of deepfake technology through the lens of information poverty theory, identifying the risks and potential for exacerbating information inequality. The specific objectives of the study were to (a) discuss the risks posed by deepfakes in contributing to information poverty, including the spread of misinformation and the widening of existing digital divides, and (b) to provide recommendations for policy-making and educational initiatives that address the deepfake-related challenges while considering the equitable distribution of information in the digital age. This review and qualitative study and approach may contribute more to the literature on the deepfake domain from a developing country perspective.

4.1. Methods

This qualitative review study investigated the phenomenon of deepfake technology and its implications through a comprehensive synthesis of secondary data. Recognising the complexity of deepfakes as a social and technological issue, the review employs a diverse range of qualitative methods in its investigative approach.

This literature review study employs a qualitative methodology to systematically gather and examine scholarly articles, technical reports, and authoritative media sources concerning deepfakes. Our goal is to dissect the dualistic nature of deepfake technology, quantifying its potential for positive applications against its capacity for malfeasance.

Utilising Scopus library databases, we curated literature, specifically journal articles and conference papers that pertained to deepfakes. Keywords such as "deepfake detection," "deepfake technology," and "misinformation" guided the database search. Additionally, a few sources were identified through cross-referencing bibliographies of pivotal studies, augmenting our database with a wider spectrum of literature.

Publications considered relevant were those discussing the technological aspects of deepfakes, their societal impact, detection methods, and policy responses. Exclusion criteria applied were to non-peer-reviewed articles, non-English publications, and those not directly related to the study's main purpose.

This investigation utilised a thematic analysis to identify emergent patterns and synthesise disparate findings across the literature. Key themes, such as technological advancements, legal and ethical implications, and misinformation dynamics, were examined. The study employed conceptual mapping to visualise connections between various aspects of deepfake technology and their societal implications. Lastly, critical appraisal tools were used to assess the quality and relevance of each selected study, aiming to ensure the reliability and validity of our conclusions.

The qualitative data garnered from the selected sources were compiled into a cohesive narrative, linking technological advances in deepfake creation and detection to their broader social and ethical implications. The literature review culminates in a critical discussion,

summarising the state-of-the-art in the deepfake technology concept, highlighting gaps in current knowledge, and proposing directions for future research.

5. Discussion

The discussion section first discusses the detection of deepfakes and the need for advanced solutions. This section introduces the problem and establishes the need for sophisticated detection methods. Secondly, the discussions focuses on deepfake detection and mitigation strategies. This section details the strategies and methods employed to detect and mitigate deepfakes. Lastly, the study discusses detecting and mitigating deepfakes through multi-faceted defences. This section summarises and underscores the importance of a comprehensive, multi-dimensional approach to address the issue effectively. This order creates a narrative flow from identifying the issue and discussing potential solutions to emphasising the complexity of a complete response.

5.1. Detecting deepfakes: the need for advanced solutions

The pervasive threat of deepfake technology necessitates advanced solutions to safeguard individuals and societal trust. Habbal et al. (2024) stress the importance of balancing trust, risk, and security in AI systems. They emphasise the need for frameworks that consider transparency, explainability, fairness, accountability, and the ethical deployment of AI to build user trust and address the vulnerabilities introduced by AI.

Advanced detection techniques powered by AI algorithms and machine learning models are at the forefront of this defence. The collaboration between researchers, technology companies, and law enforcement agencies further enhances detection capabilities and bolsters effective countermeasures against the ever-evolving landscape of deepfakes. For instance, Lee et al. (2021) highlight the need for reliable methods to distinguish between authentic images and sophisticated forgeries created by Generative Adversarial Networks.

Information poverty theory interweaves with the discourse, highlighting the necessity for a society equipped with essential skills to discern between authentic and manipulated content. As we delve into the multi-faceted defence against deepfakes, it is essential to consider the implications of deepfake user perceptions, acknowledging the need for user-friendly and human-centred deepfake technology detection methods. Future research should address issues discussed in this study, focusing on developing technology that aligns with human perceptions while being mindful of potential misuse (Kaate et al., 2024).

The complexity of the deepfake landscape extends beyond technological advancements. Regulations and policies are imperative to govern the responsible use of deepfake technology, preventing malicious intent. Understanding deepfake user perceptions becomes pivotal, guiding the responsible integration of deepfakes into design processes. This integration enhances user experiences and addresses concerns surrounding potential misuse.

The impact of deepfake technology spans diverse concerns, from identity theft to political deception, requiring a collective effort to mitigate risks. Public awareness campaigns and educational initiatives play a crucial role in cultivating a resilient defence. Advocating for media literacy and critical thinking skills empowers individuals to discern between real and manipulated content, addressing the societal implications of deepfakes (Ćufar, 2021).

The multi-pronged approach to deepfake defence extends beyond technological advancements. It necessitates legislative measures, educational initiatives, and public awareness campaigns. Ongoing research and development of advanced detection solutions must keep pace with the evolving nature of deepfakes. Holistic strategies, incorporating legislation, company policies, education, and public awareness, create a robust defence against the harmful impact of deepfakes.

The urgency of detecting and mitigating deepfakes mandates a comprehensive defence strategy. Technological advancements, legislative measures, educational initiatives, and public awareness campaigns must coalesce to effectively combat the multifaceted challenges posed by deepfake technology. Through collaborative efforts, we can navigate the intricate landscape of deepfakes, preserving the integrity of information in our digital age.

In the face of the escalating impact and potential harm wrought by deepfake technology, the imperative for advanced solutions to detect and combat deepfakes has become very important. For example, Shaheed et al. (2024) report on the detection of presentation attacks on biometric systems using deep learning techniques, while Pintelas and Pintelas (2022) investigate deep learning representations for 3D images, improving detection capabilities for both deepfakes and medical applications. Employing cutting-edge detection techniques, such as AI algorithms and machine learning models, is essential to accurately identifying and countering the evolving sophistication of deepfakes. A collaborative effort between researchers, technology companies, and law enforcement agencies is pivotal, enhancing detection capabilities and formulating effective countermeasures.

The multifaceted defence against deepfakes demands a holistic approach, incorporating insights from information poverty theory to address disparities in knowledge and awareness. Future research should focus on user-friendly, human-centred deepfake detection technology, addressing concerns raised in studies such as Kaate et al. (2024). Simultaneously, regulations and policies are needed to govern deepfake use, ensuring protection against malicious purposes. Understanding deepfake user perceptions is crucial in responsibly integrating deepfakes into the design process and user interface, utilising knowledge about people's characteristics or personas to enhance immersive experiences while mitigating potential misuse.

The implications of deepfake technology span a spectrum of concerns, including identity theft, political deception, media credibility, and societal trust (Ćufar, 2021). Society must be educated and aware of deepfake existence to counter these risks. Advocacy for media literacy and critical thinking is indispensable in empowering individuals to discern between real and manipulated content. Although research indicates a general weakness in users' ability to detect deepfakes, ongoing investment in advanced detection technologies, regulations, and media literacy promotion can work towards minimising negative impacts.

The collaboration between researchers, technology companies, and policymakers becomes imperative to develop advanced detection techniques, regulations, and education programmes. Public awareness campaigns and educational initiatives are vital components to inform individuals about deepfake technology and its potential risks. In summary, a comprehensive approach, integrating technological advancements, regulatory measures, media literacy, and public awareness, is necessary to counter the threats posed by deepfake technology effectively. By adopting this multi-pronged strategy, society can strive to minimise the harmful effects and ensure the integrity of information in the digital age.

5.2. Deepfake detection and mitigation strategies

AI techniques, particularly deep learning algorithms, play a pivotal role in detecting deepfakes. Trained on extensive datasets encompassing real and manipulated content, these algorithms discern patterns and identify anomalies within the altered media. For example, Ali et al. (2021) highlight the importance of preparing children to navigate a society increasingly influenced by AI, emphasising the role of education in fostering critical media literacy. Concurrently, Liz-Lopez et al. (2024) note the evolution of forensic techniques in multimedia data, focusing on video and audio manipulation detection, particularly through multimodal approaches that were less predominant in earlier research.

Forensic analysis, examining features like lighting, shadows, and facial movements, also aids in differentiating authentic content from manipulated deepfakes. Public awareness is equally crucial; educating individuals about the existence and potential harms of deepfakes

cultivates a critical media consumer base less susceptible to manipulation.

Concerns surrounding misinformation, fraud, and identity theft necessitate ongoing research and development in deepfake detection methods. Researchers and experts actively contribute to developing and implementing strategies to counteract the evolving threats posed by deepfake technology (Shi et al., 2023). To effectively combat the harmful effects of deepfakes, a combination of proactive and reactive measures is essential.

Proactive measures involve continually developing advanced detection algorithms and technologies capable of identifying and flagging potential deepfake content. Liu et al. (2023) highlight challenges in existing algorithms, noting issues with computational efficiency, generalisation, and interpretability. Reactive measures encompass fact checks and swift actions such as removing deepfake content from online platforms and alerting relevant authorities or affected individuals. Ongoing research is indispensable to stay ahead of the rapidly evolving techniques employed in deepfake creation.

Collaboration between researchers, industry experts, and policymakers is crucial in addressing the deepfake challenge comprehensively. By sharing knowledge, resources, and expertise, stakeholders can collectively develop more robust and effective deepfake detection tools. This collaborative effort aligns with the principles of information poverty theory, aiming to bridge gaps in knowledge and awareness. A significant limitation of current state-of-the-art deep neural networks is their substantial energy consumption and storage requirements (Tiwari et al., 2023; Chesney & Citron, 2019). As explored in this paper focusing on the application of information poverty theory to deepfake technology, the resource intensiveness of these models has important implications. Despite considerable advancements in artificial neural networks (ANNs), achieving biological neural networks' energy efficiency and online learning capabilities remains a challenge (Zheng et al., 2022). Neuromorphic computing, inspired by the human brain, offers a potential solution by designing hardware that mimics neural processes (Zheng et al., 2022). Central to this approach are spiking neural networks (SNNs), often referred to as the third generation of neural networks, which bridge the fields of machine learning and neuroscience (Zheng et al., 2022; Yamazaki et al., 2022). Unlike traditional ANNs that rely on continuous data streams, SNNs operate in continuous time, employing discrete electrical pulses, or spikes, for communication between neurons (Zheng et al., 2022). This shift towards more biologically plausible models could be crucial in addressing the escalating proliferation of deepfakes, particularly as SNNs become more powerful and accessible, potentially influencing both their creation and detection.

The success of deepfake detection and mitigation strategies hinges on a combination of technological advancements, regulatory measures, public awareness, and collaborative efforts. Protecting individuals and society from the damaging effects of deepfakes requires a comprehensive approach that integrates these strategies. By implementing such measures, we can work towards minimising the damage caused by deepfakes and safeguarding the integrity of our digital landscape.

5.3. Detecting and mitigating deepfakes: a multi-faceted defence

In light of the escalating prevalence and sophistication of deepfakes, an imperative arises to fortify our defences through comprehensive detection methods and the cultivation of critical thinking and media literacy. Technological and societal challenges require collaboration and a deep understanding of the evolving landscape. As we delve into the multifaceted defence against deepfakes, it is crucial to weave in insights from information poverty theory, acknowledging the disparities in access, interpretation, and application of information.

The landscape of deepfake creation tools is rapidly advancing, enabling the creation of realistic and convincing deepfakes by almost anyone (Firc et al., 2023). This accessibility heightens the urgency for effective detection and mitigation strategies. To counteract this evolving threat, advanced deepfake detection algorithms leveraging AI have become important. The continuous development of these algorithms positions researchers to outpace deepfake creators in the ongoing arms race, ensuring highly accurate detection of manipulated media. However, a legislative framework must complement this technological prowess to act as a deterrent and provide legal recourse against malicious use (Godulla et al., 2021).

In parallel, education and training programmes emerge as crucial components of the defence against deepfakes. Information poverty theory comes into play here, underscoring the importance of addressing disparities in essential skills, abilities, and resources related to information. By implementing educational initiatives, individuals can be equipped with the knowledge needed to identify, question, and report instances of deepfake manipulation. Public awareness campaigns further amplify these efforts, emphasising the existence and potential dangers of deepfakes.

The role of social media and online content-sharing platforms is pivotal. Strict guidelines and policies, including regular content monitoring, fact checks and AI-based detection systems, become imperative to prevent the proliferation of deepfakes. Collaboration between technology companies, governments, and researchers adds another layer to this defence. By sharing information and collaborating on research and development, a united front can stay ahead of the deepfake threat, safeguard privacy, uphold information integrity, and maintain public trust in institutions (Habbal et al. (2024).

Addressing the multifaceted challenges posed by deepfakes demands a holistic and interdisciplinary approach. Active involvement from researchers, policymakers, technology companies, and the general public is essential. Through collective efforts, we can effectively combat the adverse implications of deepfakes, embracing technological advancements, legislative measures, educational initiatives, and collaborative industry strategies to protect the fabric of our digital society.

6. The study's contribution to the existing deepfakes-related work

Fig. 1 below visually represents the relationship between deepfake technology and information poverty, the study's key findings, and offers a clear depiction of strategies to address the deepfake-related challenges illuminated by the research.

6.1. Interplay with information poverty theory

The article underscores deepfakes as a **potential threat to society** due to their deceptive and manipulative capabilities. This threat extends to information poverty, as individuals may need more resources or awareness to discern manipulated content.

Deepfakes have a negative impact on **trust, credibility, and identity**, particularly among college students. This impact will likely affect individuals who may already be vulnerable to information poverty.

The research emphasises the potential influence of deepfakes on **information and news production and sharing**. This influence can exacerbate information inequality, as those with limited resources may struggle to navigate a landscape tainted by manipulated information.

Ongoing research explores the impact of deepfakes **on society**, highlighting the evolving nature of this technology and the need for continuous understanding (Gamage et al., 2023).

While acknowledging the **positive potential** of deepfakes in entertainment and the creative industries, the study recognises the **associated risks**, including the spread of misinformation, threats to privacy and security, and erosion of public trust. These risks are particularly relevant to individuals facing information poverty.

Researchers and developers must address the challenges posed by deepfakes with a focus on creating safeguards and detection methods. This **call to action** is crucial to protect individuals, especially those vulnerable to information poverty, from the harmful effects of deepfake

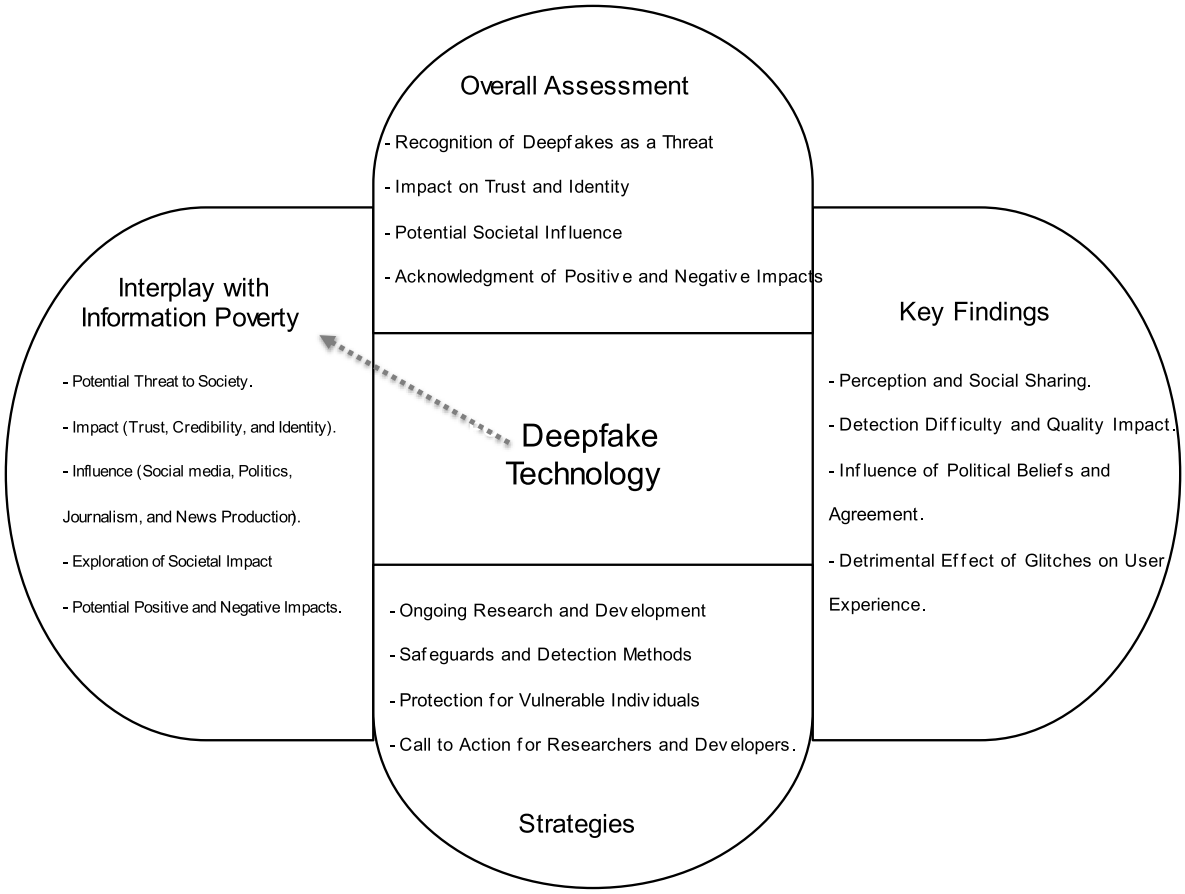


Fig. 1. Contributions to deepfakes-related work.

technology.

6.2. Key findings

Deepfakes are often **perceived as more accurate** than they are, leading to a higher likelihood of **sharing** on social media than genuine videos (Kaate et al., 2024) or information.

Users **face difficulty detecting high-quality**, realistic deepfake videos or information. The quality and believability significantly impact users' perception, trust, and engagement with the technology.

Beliefs and agreement with content influence individuals' perceptions of deepfake videos or information. Conservatives and content-agreeing viewers and users are less likely to differentiate between real and deepfake content. Perceptions of deepfake's realness impacts empathy and credibility towards the content.

Deepfake glitches can harm personas and user experiences and task performance, emphasising the importance of addressing perceptual challenges in deepfake technology implementation.

The **quality and believability** of deepfakes significantly impact users' perception, trust, and willingness to engage with the technology.

6.3. Overall assessment

Due to their deceptive and manipulative capabilities, deepfakes are seen as a **potential threat to society**. The research suggests negative impacts on trust, credibility, and identity, especially among college students (Čufar, 2021).

Deepfakes have the potential to significantly impact society in **both positive and negative ways**. Positive impacts include applications in entertainment, creative industries, and human-computer interaction. However, risks involve the spread of misinformation, threats to privacy and security, and erosion of public trust in media and information sources (Gamage, 2022).

Deepfakes' **impact on society** is an ongoing area of research, emphasising the need for continued exploration and understanding (Gamage et al., 2023).

Deepfakes pose **risks** such as the rapid spread of misinformation, threats to privacy, and political manipulation. Researchers and developers are urged to focus on creating **safeguards** and detection methods to protect individuals from the harmful effects of deepfake technology.

6.4. Strategic contributions to some deepfake detection methods

These strategies involve using various techniques and technologies to identify and combat the spread of deepfakes and some of the commonly used methods for deepfake detection are mentioned below.

Li et al. (2024) state that examining temporal identity inconsistency is essential in developing strategies to detect deepfakes. For example, Liu et al. (2023) proposed a method that analyses the consistency of an individual's facial features over time to detect deepfake videos. Individuals who lack knowledge about deepfakes may easily accept what they come across. Examining temporal identity inconsistency to detect deepfake videos directly relates to information poverty. As explained by theorists, Chatman (1996), information poverty involves individuals lacking appropriate or adequate information to make informed decisions due to a gap in knowledge or access. When Li et al. (2024) discuss methods to analyse the consistency of facial features over time, they present a way to filter out deceptive information that could, if left unchecked, contribute to information poverty. Educating individuals and equipping them with tools to recognise and understand the nature of deepfakes makes them less vulnerable to accepting falsified information

as truth. As a result, people are better equipped to distinguish between authentic and manipulated content, which enhances their information literacy and helps to prevent the spread of misinformation. Implementing such detection methods can thus be viewed as a strategic measure to address and reduce information poverty by preserving the integrity of the information that individuals receive and trust.

Choudhary and Arora (2024) advocate for the utilisation of forensic analysis methods. For instance, Choudhary and Arora (2024) suggest employing forensic analysis techniques that assess multiple video elements, including compression artifacts, lighting and shadow inconsistencies, and facial expression discrepancies to unmask deepfakes. Choudhary and Arora (2024) thus suggest using forensic analysis techniques to expose deepfake videos. Individuals who lack adequate understanding of deepfake information may be unable to perform forensic analysis of the information they come across. These methods are therefore crucial in protecting against the deepfake-related exacerbation of information poverty, ensuring that all individuals have access to reliable and verified information, thereby preventing misinformation from further disadvantaging those vulnerable to data and information manipulation.

Another method is implementing deep learning algorithms, particularly convolutional neural networks, which are widely employed in deepfake detection (Kaate et al., 2024; Gritsenko, 2024, 2024; Gritsenko, 2024; Sharmila & Aparna, 2024) These algorithms are trained on large datasets of both real and deepfake videos to develop a discriminative model that can accurately distinguish between the two. Implementing deep learning algorithms, such as convolutional neural networks for deepfake detection is indirectly linked to mitigating information poverty. The spread of deepfakes can exacerbate information poverty, as they can significantly erode the quality and reliability of information accessible to the public. Accurate deepfake detection, together with fact checks and data verification, helps ensure that individuals have access to truthful and authentic information, which is vital for making informed decisions and participating effectively in society. By improving the quality of information, deep learning algorithms contribute to combating information poverty by reducing the circulation of deceptive media that could mislead or misinform individuals.

Kaate et al. (2024) suggest utilising the multimodal analysis method. This combination of multiple modalities, such as facial expressions, voice patterns, and body movements, can greatly enhance the accuracy of deepfake detection in videos. By analysing these different modalities simultaneously, any discrepancies or inconsistencies between them can be identified, indicating the presence of a deepfake. As suggested by Kaate et al. (2024), the multimodal analysis method plays a critical role in addressing information poverty. Information poverty occurs when individuals lack access to the necessary information to understand and engage with the world around them. The presence of deepfakes can contribute to this problem by polluting the information ecosystem with false or misleading content. The method enhances the accuracy of detecting fabricated content by applying a multimodal analysis that examines inconsistencies across various indicators such as facial expressions, voice patterns, and body movements. Consequently, it aids in maintaining the integrity of the information landscape. Ensuring the circulation of authentic and reliable media is fundamental to empowering individuals with accurate knowledge, which is a key factor in alleviating information poverty.

With the increasing prevalence of deepfake technology, individuals and organisations need to be vigilant and take proactive measures to protect themselves from the potential threats posed by deepfakes (Lin et al. 2022). Linking the proactive measures against deepfake threats to information poverty theory underscores the importance of these steps in preventing a decline in informational quality and accessibility. Information poverty is characterised by a lack of essential information, limiting individuals' capacity to make informed choices and participate fully in societal and democratic processes. When individuals and organisations take vigilant actions to guard against the spread of deepfakes, they help to maintain the authenticity and reliability of the information upon which people depend. By doing so, these measures support the fight against information poverty, ensuring that all members of society have equitable access to true and useful information, which is a cornerstone of personal and community development.

7. Future research outlook

The evolving landscape of deepfake technology presents numerous avenues for future research, aimed at addressing emerging challenges and harnessing its potential responsibly. Key research directions include the development and refinement of machine learning algorithms to enhance deepfake detection accuracy, particularly in response to evolving techniques. Integrating multiple modalities, such as video and audio, could significantly improve the robustness of deepfake detection systems. This study acknowledges that the information poverty theory might not fully capture all the challenges posed by deepfakes. Future studies may need to employ several theories to address the gaps inherent in the information poverty theory in a similar study.

Collaboration between researchers, technology companies, and regulatory bodies is essential for developing standardised detection methods and frameworks and these players are a key audience for this paper. Future research should focus on establishing guidelines for responsible use, updating detection methods, and implementing legislation to address privacy infringements and defamation. Investigating the effectiveness of these collaborations can ensure a cohesive approach to tackling deepfakes.

Public awareness campaigns aimed at the general public play a crucial role in educating individuals about deepfakes and their potential dangers. Research should examine the impact and effectiveness of these campaigns in fostering critical thinking and encouraging verification of information from trustworthy sources. In addition, the ethical implications of deepfake technology in various contexts, including arts, activism, and human-computer interaction, warrant thorough investigation. Exploring the benefits of deepfakes in creative arts, storytelling, and entertainment can offer a balanced view of this technology.

Also, understanding the influence of user demographics, such as gender, age, and race, on the detection of deepfake abnormalities is vital. Research should examine the psychological and cognitive factors that contribute to users' susceptibility to deepfake manipulation. This knowledge can inform the development of targeted educational campaigns and interventions, ultimately enhancing the public's ability to discern authentic content from manipulated media.

8. Recommendations

To effectively combat the challenges posed by deepfake technologies and foster their responsible use, the following recommendations are proposed: Advocate for a comprehensive and multifaceted approach, incorporating legal measures, educational initiatives, and technological advancements to address the complex issues surrounding deepfake technology. Emphasise the importance of fostering a collaborative environment between experts, technology companies, regulatory bodies, and the general public, encouraging open dialogue and knowledge-sharing to stay ahead of evolving deepfake threats.

Continuous investment in research and technology development is crucial for enhancing deepfake detection methods. Collaboration between researchers, technology companies, and regulatory bodies should be prioritised to develop robust frameworks. Public awareness and education campaigns must be ongoing to inform individuals about the risks of deepfakes and the importance of critical thinking, empowering the public to verify information from reliable sources.

Quality and believability should be prioritised in deepfakes for user experience design. Developers must adhere to ethical guidelines for the responsible implementation of this technology. By implementing these recommendations and pursuing future research directions, society can

collectively work towards minimising the negative impacts of deepfake technology while exploring its potential benefits responsibly. Linking these aspects with information poverty theory can provide a more comprehensive understanding of the challenges and opportunities faced in the digital age.

9. Conclusion

This study has provided a nuanced examination of deepfake technology through the lens of information poverty theory. The analysis has underscored the inherent risks deepfakes pose, emphasising their potential to exacerbate information inequality. The widespread dissemination of misinformation and the deepening of digital divides are identified as significant contributors to information poverty in the digital age, especially in developing countries. Recognising the situation's urgency, this study advocates for proactive measures to mitigate the impact of deepfakes on information equity. The study reinforces the need for comprehensive policy-making and educational initiatives to address in particular the socio-economic challenges presented by deepfakes. Policies should be designed to deter malicious uses of deepfake technology, safeguard the integrity of information, and uphold democratic values. Educational programmes are crucial to empower individuals with the skills necessary to navigate an environment rife with manipulated content.

The findings of this study hold several practical implications for stakeholders in diverse domains. Policymakers can utilise the insights provided to formulate and enact legislation that specifically targets the responsible use of deepfake technology. Collaborative efforts between technology companies, regulatory bodies, and researchers are crucial for developing and implementing effective strategies, particularly in developing country contexts where information poverty is more extreme. Educational institutions can integrate content on deepfakes and media literacy into their curricula, ensuring that individuals are equipped with the knowledge and critical thinking skills required to discern between authentic and manipulated content. Technology companies and online platforms should consider implementing stringent guidelines, utilising advanced detection technologies, and actively participating in public awareness campaigns to curb the spread of deepfakes.

Despite the contributions of this study, certain limitations should be acknowledged. The rapidly evolving nature of deepfake technology means that some aspects covered in this analysis may become outdated. Additionally, the scope of the study is limited to the information available up to the knowledge cutoff date. Future developments in deepfake technology may introduce new challenges not explored in this study. The study also relies on existing literature and may be subject to the limitations inherent in the available research. The subjective nature of analysing the impact of deepfakes on information poverty introduces an element of interpretation and may vary based on individual perspectives.

Going forward, the largely qualitative assessment provided in this paper should be supported by more empirical data and research. This will enhance this paper's findings targeting policymakers, technology companies and academics so that in turn the general public can be better informed about deepfake technologies through educational and awareness programmes.

In conclusion, while this study offers valuable insights, ongoing research remains essential to stay abreast of technological advancements and evolving societal responses to deepfake challenges. In light of these challenges and the potential harm caused by deepfakes, initiatives such as the Deepfake Detection Challenge by Microsoft and the Partnership on AI have been launched to encourage collaboration and innovation in developing effective deepfake detection technologies. These initiatives aim to foster the development of advanced algorithms, datasets, and evaluation frameworks to enhance the detection capabilities of both human reviewers and automated systems. The goal is to enable swift and accurate identification of deepfakes, ensuring that individuals, organisations, and society can navigate the digital landscape confidently and trust in the authenticity of the content they encounter.

CRediT authorship contribution statement

Walter Matli: Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jjimei.2024.100286.

References

Ali, S., DiPaola, D., Lee, I., Sindato, V., Kim, G., Blumofe, R., & Breazeal, C. (2021). Children as creators, thinkers and citizens in an AI-driven future. *Computers and Education: Artificial Intelligence, 2*, Article 100040. https://doi.org/10.1016/j.caeai.2021.100040

Anwar, M. A., Tahir, S. F., Fahad, L. G., & Kifayat, K. (2023). Image forgery detection by transforming local descriptors into deep-derived features. *Applied Soft Computing, 147*, Article 110730. https://doi.org/10.1016/j.asoc.2023.110730

Ahmed, S. (2021). Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes. Personality and Individual Differences, 182, p.111074.

Bonomi, M., Pasquini, C., & Boato, G. (2021). Dynamic texture analysis for detecting fake faces in video sequences. *Journal of Visual Communication and Image Representation, 79*, Article 103239. https://doi.org/10.1016/j.jvcir.2021.103239

Borji, A. (2022). Pros and cons of GAN evaluation measures: New developments. *Computer Vision and Image Understanding, 215*, Article 103329. https://doi.org/10.1016/j.cviu.2021.103329

Britz, J. J. (2004). To know or not to know: A moral reflection on information poverty. *Journal of Information Science, 30*(3), 192–204.

Buo, S.A. (2020). The emerging threats of deepfake attacks and countermeasures. arXiv preprint.

Byeon, H., Shabaz, M., Shrivastava, K., Joshi, A., Keshta, I., Oak, R., Singh, P., & Soni, M. (2024). Deep learning model to detect deceptive generative adversarial network generated images using multimedia forensics. *Computers & Electrical Engineering, 113*, Article 109024. https://doi.org/10.1016/j.compeleceng.2023.109024. -109024.

Chatman, E. A. (1996). The impoverished life-world of outsiders. *Journal of the American Society for Information Science, 47*(3), 193–206.

Chen, P., Liu, J., Tao, L., Cai, Y., Zou, S., Dai, J., & Han, J. (2021). DLFMNet: End-to-End detection and localization of face manipulation using multi-domain features. https://scite.ai/reports/10.1109/icme51207.2021.9428450.

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., 107, p.1753.

Choudhary, A., & Arora, A. (2024). Assessment of bidirectional transformer encoder model and attention based bidirectional LSTM language models for fake news detection. *Journal of Retailing and Consumer Services, 76*, Article 103545. https://doi.org/10.1016/j.jretconser.2023.103545. -103545.

Čufar, K. (2021). Legal aspects of content moderation on social networks in Slovenia. https://scite.ai/reports/10.54237/profnet.2021.mwsm_6.

Dasilva, J.P., Ayerdi, K.M., & Galdospin, T.M. (2021). Deepfakes on twitter: Which actors control their spread? https://scite.ai/reports/10.17645/mac.v9i1.3433.

Firc, A., Malinka, K., & Hanáček, P. (2023). Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors. *Heliyon, 9*(4), e15090. https://doi.org/10.1016/j.heliyon.2023.e15090. -e15090.

Freeman, C., & Louçã, F. (2001). *As time goes by: from the industrial revolutions to the information revolution*. Oxford University Press.

Fu, X., Li, S., Yuan, Y., Li, B., & Li, X. (2023). Forgery face detection via adaptive learning from multiple experts. *Neurocomputing, 527*, 110–118. https://doi.org/10.1016/j.neucom.2023.01.017

Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M.E., & Sasahara, K. (2022, April). Are deepfakes concerning? Analyzing conversations of deepfakes on Reddit and exploring societal implications. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1–19).

Gamage, D., Ravinthran, H., & Sasahara, K. (2023). Moral intuitions behind deepfake-related discussions in Reddit communities. arXiv preprint.

Godulla, A., Hoffmann, C., & Seibert, D. (2021). Dealing with deepfakes – an interdisciplinary examination of the state of research and implications for communication studies. https://scite.ai/reports/10.5771/2192-4007-2021-1-72.

Gong, D. (2020). Deepfake Forensics, an AI-synthesized Detection with Deep Convolutional Generative Adversarial Networks. https://scite.ai/reports/10.30534/ijatcse/2020/58932020.

Goulding, A. (2001). Information poverty or overload? *Journal of Librarianship and Information Science, 33*(3), 109–111.

Gritsenko, D. (2024). Advancing UN digital cooperation: Lessons from environmental policy and governance. *World Development, 173*, Article 106392. https://doi.org/10.1016/j.worlddev.2023.106392. -106392.

Guo, Z., Yang, G., Chen, J., & Sun, X. (2021). Fake face detection via adaptive manipulation traces extraction network. *Computer Vision and Image Understanding, 204*, Article 103170. https://doi.org/10.1016/j.cviu.2021.103170

Guo, Z., Yang, G., Wang, D., & Zhang, D. (2023). A data augmentation framework by mining structured features for fake face image detection. *Computer Vision and Image Understanding, 226*, Article 103587. https://doi.org/10.1016/j.cviu.2022.103587. -103587.

Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial Intelligence Trust, Risk and Security Management (AI TRiSM): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications, 240*, Article 122442.

Hamed, S. K., Ab Aziz, M. J., & Yaakub, M. R (2023). *A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion.* Heliyon.

Harbinja, E., Edwards, L., & McVey, M. (2023). Governing ghostbots. *Computer Law & Security Review, 48*, Article 105791. https://doi.org/10.1016/j.clsr.2023.105791

Hosler, B., Salvi, D., Murray, A. F., Antonacci, F., Bestagini, P., Tubaro, S., & Stamm, M C. 2021. Do Deepfakes feel emotions? A semantic approach to detecting deepfakes via emotional inconsistencies. https://scite.ai/reports/10.1109/cvprw53098.2021.00112.

Jeong, Y., Kim, D., Ro, Y., & Choi, J. (2022). FrePGAN: Robust Deepfake Detection Using Frequency-Level Perturbations. https://scite.ai/reports/10.1609/aaai.v36i1.19990.

Kaate, I., Salminen, J., Santos, J. M., Jung, S. G., Almerekhi, H., & Jansen, B. J. (2024). There is something rotten in Denmark": Investigating the Deepfake persona perceptions and their implications for human-centered AI. *Computers in Human Behavior Artificial Humans, 2*(1), Article 100031. https://doi.org/10.1016/j.chbah.2023.100031. -100031.

Kanwal, S., Tehsin, S., & Saif, S. (2022). Exposing AI generated Deepfake images using siamese network with triplet loss. https://scite.ai/reports/10.31577/cai_2022_6_1541.

Khormali, A. and Yuan, J. 2022. DFDT: An end-to-end Deepfake detection framework using vision transformer. https://scite.ai/reports/10.3390/app12062953.

Lee, S., Tariq, S., Shin, Y., & Woo, S. S. (2021). Detecting handcrafted facial image manipulations and GAN-generated facial images using Shallow-FakeFaceNet. *Applied soft computing, 105*, Article 107256. https://doi.org/10.1016/j.asoc.2021.107256

Li, Q., Gao, M., Zhang, G., Zhai, W., Chen, J., & Jeon, G. (2024). Towards multimodal disinformation detection by vision-language knowledge interaction. *Information Fusion, 102*, Article 102037. https://doi.org/10.1016/j.inffus.2023.102037. -102037.

Liang, P., Liu, G., Xiong, Z., Fan, H., Zhu, H., & Zhang, X. (2023). A facial geometry-based detection model for face manipulation using CNN-LSTM architecture. *Information Sciences, 633*, 370–383. https://doi.org/10.1016/j.ins.2023.03.079

Lin, C., Deng, J., Hu, P., Shen, C., Wang, Q., & Li, Q. (2022). Towards benchmarking and evaluating Deepfake detection. https://scite.ai/reports/10.48550/arxiv.2203.02115.

Lingel, J., & Boyd, D. (2013). Keep it secret, keep it safe": Information poverty, information norms, and stigma. *Journal of the American Society for Information Science and Technology, 64*(5), 981–991.

Liu, X., Yu, Y., Li, X., & Zhao, Y. (2023). Magnifying multimodal forgery clues for deepfake detection. *Signal Processing: Image Communication, 118*, Article 117010. https://doi.org/10.1016/j.image.2023.117010. -117010.

Liz-Lopez, H., Keita, M., Taleb-Ahmed, A., Hadid, A., Huertas-Tato, J., & Camacho, D. (2024). Generation and detection of manipulated multimodal audiovisual content: Advances, trends and open challenges. *Information Fusion, 103*, Article 102103. https://doi.org/10.1016/j.inffus.2023.102103

Lyu, Siwei. "DeepFake Detection: Current Challenges and Next Steps." arXiv preprint , 2020.

Mijwil, M. M., Doshi, R., Hiran, K. K., Unogwu, O. J., & Bala, I. (2023). MobileNetV1-based deep learning model for accurate brain tumor classification. *Mesopotamian Journal of Computer Science, 2023*, 29–38.

Miranda-García, A., Rego, A. Z., Pastor-López, I., Urquijo, B. S., Tellaeche, A., Gaviria, J. O., & Bringas, P. G (2024). Deep learning applications on cybersecurity: A practical approach. *Neurocomputing, 563*, Article 126904. https://doi.org/10.1016/j.neucom.2023.126904. -126904.

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q. V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding, 223*, Article 103525.

Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital society, 1*(2), 19.

Pintelas, E., & Pintelas, P. (2022). A 3D-CAE-CNN model for deep representation learning of 3D images. *Engineering Applications of Artificial Intelligence, 113*, Article 104978. https://doi.org/10.1016/j.engappai.2022.104978

Qiu, H., Chen, S., Gan, B., Wang, K., Shi, H., Shao, J., & Liu, Z. (2023). Few-shot forgery detection via guided adversarial interpolation. *Pattern Recognition, 144*, Article 109863. https://doi.org/10.1016/j.patcog.2023.109863

Raj, S., Mathew, J., & Mondal, A. (2023). FDT: A python toolkit for fake image and video detection. *SoftwareX, 22*, Article 101395. https://doi.org/10.1016/j.softx.2023.101395. -101395.

Reviglio, U. (2022). The Algorithmic Public Opinion: a Policy Overview. https://scite.ai/reports/10.31235/osf.io/bjfkm.

Seow, J., Lim, M., Phan, R. C., & Liu, J. K. (2022). A comprehensive overview of deepfake: Generation, detection, datasets, and opportunities. *Neurocomputing, 513*, 351–371. https://doi.org/10.1016/j.neucom.2022.09.135

Shaheed, K., Szczuko, P., Kumar, M., Qureshi, I., Abbas, Q., & Ullah, I. (2024). Deep learning techniques for biometric security: A systematic review of presentation attack detection systems. *Engineering Applications of Artificial Intelligence, 129*, Article 107569. https://doi.org/10.1016/j.engappai.2023.107569. -107569.

Sharmila, S., & Aparna, C. (2024). Tracing footprints of anti-forensics and assuring secured data transmission in the cloud using an effective ECCDH and Kalman Filter. *Journal of Network and Computer Applications, 221*, Article 103762. https://doi.org/10.1016/j.jnca.2023.103762. -103762.

Shi, L., Zhang, J., & Shan, S. (2023). Real Face Foundation Representation Learning for Generalized Deepfake Detection. https://scite.ai/reports/10.48550/arxiv.2303.08439.

Sweetland, J.H. (1993). Information poverty—Let me count the ways. Database, 16(4), 8–10.

Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., & Maida, A. (2019). Deep learning in spiking neural networks. *Neural networks, 111*, 47–63.

Tiwari, A., Dave, R., & Vanamala, M. (2023). Leveraging deep learning approaches for deepfake detection: A review. In *Proceedings of the 2023 7th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence* (pp. 12–19).

Varsha, P. S., Chakraborty, A., & Kar, A. K. (2024). How to undertake an impactful literature review: Understanding review approaches and guidelines for high-impact systematic literature reviews. *South Asian Journal of Business and Management Cases, 13*(1), 18–35. https://doi.org/10.1177/22779779241227654

Wang, G., Jiang, Q., Jin, X., Li, W., & Cui, X. (2022). MC-LCR: Multimodal contrastive classification by locally correlated representations for effective face forgery detection. *Knowledge Based Systems, 250*, Article 109114. https://doi.org/10.1016/j.knosys.2022.109114. -109114.

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. https://scite.ai/reports/10.22215/timreview/1282.

Whyte, C. (2020). Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. *Journal of cyber policy, 5*(2), 199–217.

Whittaker, L., Kietzmann, J., Letheren, K., Mulcahy, R., & Russell-Bennett, R (2023). Brace yourself! Why managers should adopt a synthetic media incident response playbook in an age of falsity and synthetic media. *Business Horizons, 66*(2), 277–290. https://doi.org/10.1016/j.bushor.2022.07.004

Yadav, A., & Vishwakarma, D. K. (2023). MRT-Net: Auto-adaptive weighting of manipulation residuals and texture clues for face manipulation detection. *Expert Systems with Applications, 232*, Article 120898. https://doi.org/10.1016/j.eswa.2023.120898. -120898.

Yang, H., Rahmanti, A.R., Huang, C., & Li, Y. (2022). How Can Research on Artificial Empathy Be Enhanced by Applying Deepfakes? https://scite.ai/reports/10.2196/29506.

Yu, L. (2006). Understanding information inequality: Making sense of the literature of the information and digital divides. *Journal of Librarianship and Information Science, 38*(4), 229–252.

Yungui, C., Chen, J., Huang, L., Huang, T., & Ye, F. (2023). Three-classification face manipulation detection using attention-based feature decomposition. *Computers & Security, 125*, Article 103024. https://doi.org/10.1016/j.cose.2022.103024. -103024.

Yamazaki, K., Vo-Ho, V. K., Bulsara, D., & Le, N. (2022). Spiking neural networks and their applications: A review. *Brain Sciences, 12*(7), 863.

Zheng, S., Qian, L., Li, P., He, C., Qin, X., & Li, X. (2022). An introductory review of spiking neural network and artificial neural network: from biological intelligence to artificial intelligence. arXiv preprint arXiv:2204.07519.