Human-centric Computing and Information Sciences

July 2025 | Volume 15

KCIA
Korea Computer
Industry Association

www.hcisjournai.com

HCIS Human-centric Computing and Information Sciences

RESEARCH (Research Manuscript) Open Access

Human-centric Computing and Information Sciences (2025) 15:38
DOI: https://doi.org/10.22967/HCIS.2025.15.038
Received: March 20, 2024; Accepted: April 28, 2024; Published: July 15, 2025

Cybersecurity, Digital Forensics, and the IoT for Deepfake Investigation on Social Media Platforms: A Review

Abdullah Ayub Khan[1,*], Asif Ali Laghari[2], Rex Bacarra[3], Roobaea Alroobaea[4], Sultan Algarni[5], Abdullah M. Baqasah[6], and Jamil Abedalrahim Jamil Alsayaydeh[7,*]

Abstract

Numerous privacy and security concerns could jeopardize the social media ecosystem: mishandled hotspot connectivity used as a standalone cybercrime; illegal campaigning; deviant activities conducted online; propagation of anti-national propaganda; and recruitment of terrorists through social media platforms. These provide significant challenges for social media firms and open up new avenues for digital forensics investigations to support analysis and provide answers in the ever-evolving realm of cyber security. However, in terms of logging, inspecting, assessing, storing, presenting, and recording deepfake content, the growth of the Internet of Things has an impact on the cyber forensics' investigation lifecycle. The integration of IoT with the most recent cyber security frameworks for deepfake analysis in social media platforms is reviewed in detail in this study. We also examine deepfake sounds, movies, and images by reviewing the media forensics literature. We combine the manipulation and alteration that takes place during social media posting for the data modality with an assessment of the technology developments in identifying and measuring manipulation, forgeries, and alterations. In conclusion, this study poses several significant research problems that should be considered, including upholding a safe chain of custody, implementing impersonation countermeasures, and fending against deceptive social engineering tactics.

Keywords

Internet of Things (IoT), Cyber security, Artificial Intelligence (AI), Blockchain, Digital Forensics, Social media, Deepfake

1. Introduction

***Corresponding Author:** Jamil Abedalrahim Jamil Alsayaydeh (jamil@utem.edu.my), Abdullah Ayub Khan (abdullah.khan00763@gmail.com)
[1]Department of Computer Science, Bahria University Karachi Campus, Karachi, Pakistan
[2]Software Collage, Shenyang Normal University, Shenyang, China
[3]Department of General Education and Foundation, Rabdan Academy, Abu Dhabi, United Arab Emirates
[4]Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia
[5]Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia.
[6]Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia
[7]Department of Engineering Technology, Fakulti Teknologi Dan Kejuruteraan Elektronik Dan Komputer (FTKEK), Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia

The last several years have seen a rapid development in the fields of artificial intelligence (AI), blockchain technology, big data, and the Internet of Things (IoT). This has led to the creation of new paradigms and the development of tools and techniques for use in multimedia investigation [1, 2]. However, advancements in deep learning (DL) and machine learning (ML) have registered a variety of algorithms for use in multimedia manipulation, which goes beyond the realm of tampering and forgery.

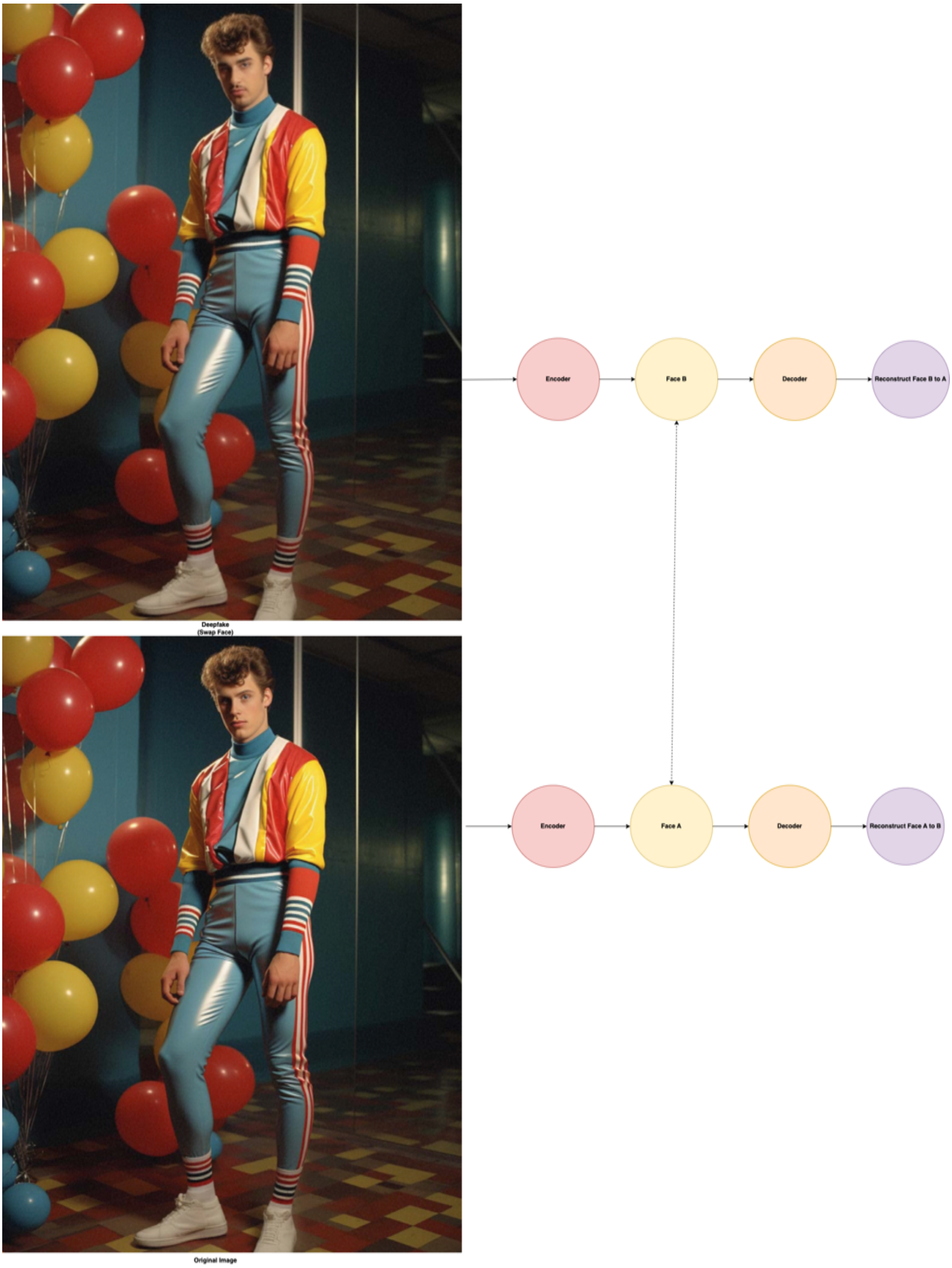

Fig. 1. Working operation of deepfake generation and the role of discriminator.

Notably, these technological advancements have been extensively employed for amusement reasons on social media platforms, but they have also been widely used in genuine platforms, such as the creation of a Metaverse environment [1–3]. As technology evolves, malevolent users take advantage of them to carry out evil and illegal deeds. The following are some of the behaviors listed: (1) disseminating the hate of well-known politicians, harassing people via email, manipulating images, (2) high-definition (HD) false video streaming, and (3) picture alteration. Remarkably, when compared to the classical AI, the advanced AI has better manipulation abilities for content linked to images, audio, and video. FakeApp, FaceApp, image enhancer, and other popular open-source software programs are registered as facial image and video manipulators because they produce a realistic-looking view [4, 5]. In this way, the public can exchange facial aspects, including appearance, hairdo, age hierarchy, and personal characteristics, using these manipulating platforms. However, the spread of this ideology of fabricating information leads to a number of harassment- and cognitive-related issues, including bullying.

Realistic photos and HD video streaming have been regarded as the current state of the art in deepfake due to the potential for multimodal manipulation [2, 3]. The phrase originated from the 2017 approach developed by Reddi, an unnamed company that combined multimedia technologies and deep learning algorithms to generate photo-realistic false films by unethically swapping the faces of two people. Several approaches are offered that explain how to assess the issues that arise in real-time video streaming in order to counter such multimedia deepfakes. A face swapping network that is currently utilized to create fraudulent photos and videos utilizing an encoder and a decoder is depicted in Fig. 1 as the working mechanism to develop this type of counterfeit multimedia content: a generator and a discriminator. The majority of applications are created solely for entertainment, but in the most recent instance, we noticed something unusual. Furthermore, the discriminator cycle is identified as the network possessing the genuine ability to produce novel images. In order to accomplish this, a technique known as the generative adversarial network (GAN) is developed through the cooperation of two neural nets, including a generator and a discriminator [6, 7]. On the other hand, a number of other techniques have been put up for manipulating and switching between multimedia information. The effectiveness of each of these methods in terms of their detection capacity when used on various social media platform datasets, including Facebook, Instagram, and Twitter. The deepfake investigation cycle is crucial, and none of the Al-enabled ML and DL techniques work well enough to identify deepfakes in multimedia content without going against the relevant associations and the chain of custody hierarchy [3, 5, 6]. In terms of choosing legal action, it is beginning to look into more.

The yearly report on multimedia forensics investigation states that generative networks have allowed AI specialists to try various deepfake-related breakthroughs. For example, the situation of facial re-enactments is improved by the developments in computer vision [7, 8]. The well-known suggestion in the resume is as follows: Face2Face, ReFace, and DeepSwap are the first three. The primary function of the aforementioned applications is to transfer facial expressions in real-time from an actual image to a digital avatar. A cycle-GAN was offered as an improvement from AI, particularly in DL, that aids in the transformation of images and videos by adding styles in a unique way [8, 9]. Researchers from Barkley University and the University of Washington, in particular, reported simulation results based on lip movement synchronization when speaking about the video content that originates from several sources [9, 10]. But when the first illegal activity—sharing pornographic movies of celebrities in which the face is specifically replaced with the genuine object—occurs, the idea of deepfake is brought forward. Conversely, a service called "deepfake creation" has been introduced, encouraging users to access fake content investigation processes through various websites. This service is supported by private sponsors, particularly through social media channels. The deepfake creation services were removed from several platforms, including social media accounts, a while back (i.e., Twitter).

Numerous deepfake countermeasures are prioritized in light of the possible threats and hazards associated with social media privacy vulnerabilities. An approach that aids in the training of extensive data collection—limited to social media data—was presented by the author of [10, 11]. Face forensics, a subset of social media forensics (SMF), was used in this way to investigate Deepfake content. The identification of deep video portraits—which are intended to facilitate the analysis of photo-realistic re-animation of multimedia-based videos—is made possible by these kinds of advancements. Furthermore, the development of AI techniques, like the derivation of GAN, modifies the dynamics by expanding the use of deep-face films to deepfake body movement, which is the transfer of a human body movement to another human [11, 12]. In this overall scenario, NVIDIA contributes significantly to the style-based generator process, where the architecture is made to support generations of synthetic multimedia [11–13]. The report, which was produced by Google at the end of 2022 and the beginning of 2023, makes it very evident which of the several websites that are currently distributing deepfake content [9, 11, 13, 14]. The following is a mention of the list of discussions:

- Over 6,000 websites are under investigation and have restrictions placed on them.
- There are about 9 million deepfake videos registered for amusement, but they are most likely used for harassment and intimidation.
- As a result, just one-fourth of the social media channels can be identified and eliminated.

In the current context of multimedia research, the IoT has a measurable role. This is because AI approaches related to IoT have advanced significantly, but it also covers some difficult prospects that are crucial for quality-of-results (QoR) but are hard to accomplish with current methods. In order to estimate computing cost, energy consumption, and network bandwidth, such as throughput and latency, the majority of these QoRs evaluations require model integrations, such as convolutional neural network (CNN) with inference accuracy and quality-of-services (QoS) [15, 16]. Lately, nevertheless, practically all AI models have been independently constructed and developed in compliance with IoT device protocols. It results in effective and dependable solutions such as the cataloging of significant class items, partial data gathering, object framing based on the data obtained, and categorization. For example, the effect of such collaboration represents the notable data members independently, which aids in the appearance of tracking data progress. This method, which is more effective than the traditional ones, detects deepfakes on IoT devices that mostly have minimal memory requirements by utilizing lightweight AI-enabled ML/DL algorithms. The process for detecting deepfake involves two steps: initially, the DL technique CNN is made to gather multimedia content from various social media platforms and use it in conjunction with other tools to conduct facial simulation tests. In order to improve classification, it is necessary to detect deepfakes in multimedia materials and to reduce the influence of backdrop before deleting it. Second, by actively identifying phony elements in live multimedia, IoT devices capture and alter live multimedia broadcasts. IoT devices' minimal processing overhead inherently recognizes phony prototypes that are active, particularly in social media settings.

Cyber security (CS) has evolved to become a breakthrough in Advance Digital Technology (ADT); its function is to protect the digital environment from the threat of deepfake attacks. The computer science community uses the potential applications of AI, ML, and DL to propel new paradigms. The aforementioned technologies, which were originally employed by cybercriminals to create deepfake content, are now most commonly used to identify and counteract vice versa [16, 17]. The frequency of these increasingly complex deepfakes is rising in tandem with the accessibility of technology, which has significant security consequences. Since the technology may alter physical characteristics, it can be studied in cyberspace using price matching, voice cancellation, and partial facial feature analysis. Replicating the majority of the traits is not difficult. As a result, there is an opportunity to review and evaluate as indicated. The professionals in digital forensics (DF) have faced considerable challenges in identifying and analyzing evidence that has been manipulated in order to put an end to the increasing number of deepfakes. Multimedia chain of custody is one method that several technological experts suggested to build procedures for the assessment of multimedia content, particularly in social media platforms, in order to counteract the growing threat of deepfake [15–17].

1.1 Motivation, Objectives, and Contributions

Three subcategories, including research, comprehend, and implications, make up the review process.

In light of this scenario, we create a review plan in which this document notes the necessity of identifying and choosing previously published state-of-the-art publications and explains why they should be included in the proposed study. Create a standard, a process, and an assessment as well to help with the review planning elements. Alternatively, carrying out the review is an additional project for this task that aids in the study's separation of principles, directs their examination, and presents the strategies that must be used for the investigation of multimedia deepfakes. Nonetheless, the following research questions—which are illustrated in Fig. 2—are highlighted by this work:

- To find relevant studies that are needed to complete this review.
- To identify a collection of research questions and a search strategy that complements the context of multimedia-enabled deepfakes investigation (e.g., characterizing datasets, selecting common characteristics, investigating models, and determining measurement measures to assess system performance).
- To offer a criterion for selection (from Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, etc.) in which the suggested work satisfies the requirements for quality evaluation by guaranteeing an impartial and pertinent selection of research articles during the discovery process.
- The selection of publications spans around 10 years, from 2014 to 2023.

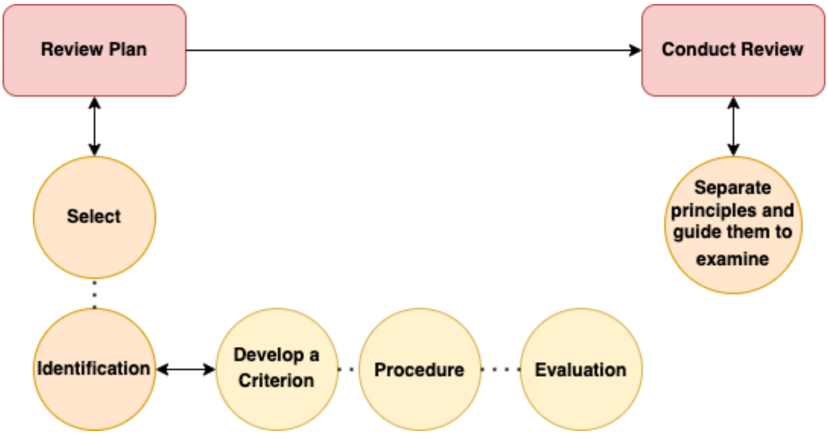

Fig. 2. Strategy of collecting articles from different sources.

One hundred thirty research articles that are indented to support the suggested research questions serve as the basis for a summary of the search technique. As we gather research on deepfake in social media, we take into account all possible search combinations based on relevant terms, without any prejudice. The refined search criterion is entirely dependent on Boolean logic, utilizing the operations "AND (+)," "OR (|)," and "NOT (!)." "Deepfake + social media platform" is the general outline of primary searching, followed by "The role of digital forensics investigation | media forensics + multimedia deepfakes! survey." Furthermore, this work's search strategy goes beyond using just one or two sources. In essence, the repositories of prospects that are being searched are as follows (Fig. 3): Web of Science (WoS), ScienceDirect, Springer Link, IEEE Xplore, ACM Digital Library, Semantic Net, and Google Scholar.

- The following highlights this review paper's main goals and contributions:
- The report evaluates more than a hundred research publications to identify recently published paradigms regarding the effects of deepfakes and countermeasures, especially in the context of social media. This highlights the relevance of media forensics. The study's goal is to present the list of deepfake investigation frameworks that a multimedia forensics specialist developed to proactively identify fake content on social media platforms.
- This paper explores the significance of technological integration, emphasizing how the convergence of cyber security, IoT, and AI enhances deepfake investigation's effectiveness, efficiency, speed, and dependability.

- A systematic examination between previously published classical strategies with the modernized ones, which presents different kinds of opportunities and drawbacks available in the current architectural updates.
- Finally, various open research prospects are addressed, along with the detailed hierarchy of how they will be addressed in the future.

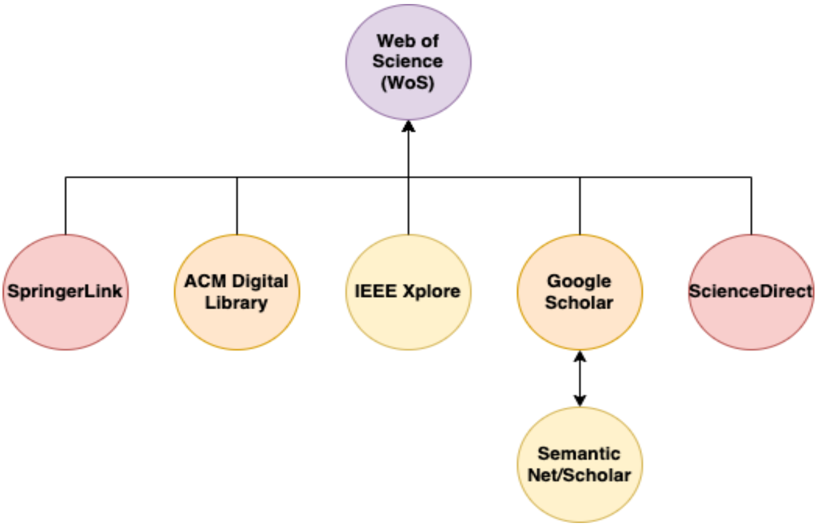

Fig. 3. Strategy of collecting articles from different sources.

1.2 Outline of this Research Article

This review article's remaining content is organized and optimized as follows. The list of social media advancements is discussed in Section 2 along with the variables that propel this technology to the next level for the coming generation. Section 3 discusses the use of IoT in social media research. This section also goes into further detail about the IoT context of deepfake and the steps taken to start taking precautions in the social media environment. In Section 4, cyber security is integrated into social media to safeguard multimedia content, their hierarchy, and the entire infrastructure. Section 5 examines the effects of working together on the concepts of DF and incident response to implement an effective approach for investigating multimedia content on social media platforms. The statement of conclusion and the discussion of outstanding research questions and future directions in Section 6 bring this study to a close.

2. Social Media Forensics

SMF technology includes several DF and cyber investigation platforms for analysis [18, 19]. This includes material extraction from social media apps such as Facebook, LinkedIn, WhatsApp, Twitter, and Instagram examining and archiving these gathered materials that are freely accessible through various media outlets in order to identify deepfake investigations. However, it can be challenging to start an investigation into similar content in other situations. Advances in the lifecycle are necessary because deeper data must be gathered for in-depth transformation and subcategorized analysis [18]. The Cyber Investigation Bureau's study states that following the proceedings to the years 2015–2022, law enforcement counters social media occurrences side by side, increasing the leverage by up to 70% [18–20].

As an illustration of SMF in action, consider the Detroit Police Department's investigation of a crime through the use of a social media post in which a suspect was seen dancing in order to deceive law enforcement into four separate attempts at robbery and subsequent harassment. SMF is involved in all of this, but it is particularly crucial in earlier situations, such as evaluating the Police department by reporting missing person cases to them. All of this can occur when utilizing AI technology for feature extraction and classification to separate the active social media handles [19, 21]. For instance, in [20, 21], a young guy went missing, and his family filed a missing person's report, claiming that the man told them to go on tour. It is one of the unfortunate victims of a car accident near Virginia that claimed the lives of many other passengers. Eventually, a composite sketch of him from the police department circulated, and at the end of 2012, someone clicked on this autopsy aid on his social media account. The agency called the family over social media to request identification after discovering and verifying dead bodies. These 17-year cases come to a finish in this way.

2.1 Process of Deepfake Investigation

Adherence to the laws that are intended to control the department, government, and legal system is the primary obligation of law enforcement authorities [20, 21]. When starting a multi-platform inquiry, it involves gathering evidence, particularly in the case of deepfake on social media. [22]. Nevertheless, a warrant is necessary in order to access audiovisual content from social networking applications due to compliance requirements with relevant agencies, police, and jurisdictions. This includes account logs, which the account holder is not allowed to upload publicly or permit any kind of accessibility. Conversely, different social media accounts can fall under different sets of rules and guidelines, which could affect the course of the inquiry. Therefore, it takes longer to be able to start gathering information when part of the data can be manipulated.

Social media investigators can use a range of methods and instruments to gather and analyze multimedia content from social media platforms in order to handle this kind of prospect [21, 22]: (1) manually navigating through accounts; (2) printing screenshots; (3) downloading copies of posted multimedia-enabled images and videos; (4) utilizing open-source platforms to preserve content; and (5) subcategorizing the hierarchy in order to create reports that can be submitted to a court of law. But, there are specific requirements for importing the SMF tools' capacity, which are crucial when it comes to using them to share content that users upload while engaging in illegal activity on social media platforms. In order to conduct an investigation more thoroughly, the first step is to choose an effective instrument that can reliably and efficiently collect and review the chain of evidence (CoE) and create an ideal chain of custody (CoC) for additional research. Benefits from an active search to the preservation of multimedia material, including social media networks, which is regarded as crucial, are offered by this hierarchy.

Table 1. A systematic review of previously published state-of-the-art frameworks

Study	List of social media deepfake scenarios	Lacks governance regulatory and protocols discussion	Possible scenario to provide solutions to regulatory and compliance	Role of social media forensics	System's integration and updates requires for design real-time deepfake on social media platforms	Detail argument based on requirement fulfillment
Kumar and Sharma [23]	D	P	N	- Capturing multimedia information - Examination - Filtration and separation - Aggregation - Analysis - Presentation, preservation, and documentation	Presented with partially discussion	√
Kwok and Koh [24]	D	D	D		Do not presented	N/A
Fagni et al. [25]	P	N	D		Presented with partially discussion	√
Yadlin-Segal and Oppenheim [26]	N	D	D		Do not presented	√
Ahmed [27]	P	D	N		Do not presented	N/A
Hancock and Bailenson [28]	N	P	P		Presented with partially discussion	N/A

N = no discussion added, P = provide in-depth description, D = partially discussed.

Based on previously published state-of-the-art studies [23–28] (Table 1), gives a systematic review report whose metrics of analysis are discussed as follows: (1) list of discussions included on the topic of social media deepfake-based scenarios, (2) lacks governance regulations and protocols related discussion, (3) possible scenario to provide solutions to regulatory and compliances, (4) role of SMF throughout the investigation, (5) systems' integration and updates requires, and (6) detail argument based on requirement fulfillment.

Table 2 presents the exploring report of deepfake investigational analysis that associated with SMF (year 2013–2023) [29–41], whose metrics of evaluation are highlighted as follows: (1) large-scale facial data, (2) face swapping, (3) application of SMF for initiate investigation process, (4) technological integration, (5) body-gesture analysis, (6) deepfake TIMIT dataset, (7) video detection modeling, (8) audio detection modeling, (9) photo detection modeling, (10) visualization, (11) chain of evidence, (12) HD multimedia content, and (13) chain of custody.

Table 2. A systematic analysis of deepfake investigation and the role of social media forensics (2013–2023)

Context of discussion	Year & study										
	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36, 37]	[38–40]	[41]	This work
Large-scale facial data	√				√		√		√		√
Face swapping		√		√		√		√			√
HD multimedia content	√	√		√			√	√	√		√
Application of SMF investigation procedure			√	√		√	√			√	√
Body-gesture analysis		√			√		√	√	√	√	√
Video detection modelling			√	√		√	√		√		√
Audio detection modelling		√			√			√		√	√
Photo detection modelling			√		√		√		√		√
Visualization	√	√		√		√		√	√	√	√
Chain of evidence										√	√
Chain of custody		√	√		√	√		√		√	√

3. Internet of Things for Social Media Forensics

A new paradigm in multimedia tampering, swapping, manipulation, and forgery has emerged as a result of the enormous advancements in AI, particularly in DL [42–46]. One of these derivations is called deepfake, in which DL uses feature transformation to produce swap pictures based on information obtained from other sources in a set-like fashion. Such evolution birth GAN offers an advanced method for creating deepfakes. Using DL techniques, it creates highly available genuine images and applies deepfake through image-to-image transformation. In this case, creating, translating, and storing deepfake content based on GAN generation requires a very high memory capacity. To successfully deploy on IoT-enabled platforms, memory efficient lightweight DL-based deepfake detection algorithms are essential for stability [42–44]. As illustrated in Fig. 4, it works in tandem with the detection application programming interface (API) to determine the most effective way to detect very advanced deepfake multimedia content generated by GANs at the edge environment. This effort certainly accomplishes a great deal of accuracy and efficiency in a short amount of time. However, SMF plays a vital role in an IoT environment by offering a new paradigm that facilitates more effective digital investigation and incident response, particularly with regard to the inspection and analysis of multimedia material in the social media domains [45, 46]. The IoT industry's technological advancements known as "digital forensics approaches" allow for a broader scope of social media inquiry through the use of correct CoE creation and deepfake detection in submitted multimedia content. IoT-Forensics' primary function is to retrieve identified data straight from social media networks.

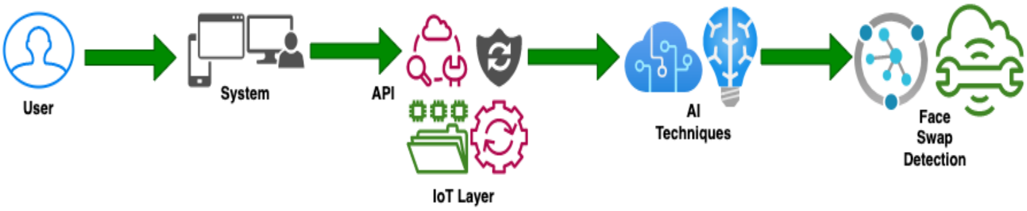

Fig. 4. Steps of IoT for SMF.

To support IoT-Forensics, however, additional activities related to multimedia information are also collected, including network traffic, cloud access logs, device registrations, and updates to personal information. These can be extracted and utilized as an extra CoE, which can then be submitted and presented in court for a decision [47–49]. The first step in the lifetime of IoT-Forensics is to conduct the tweaking of DF standard guidelines in order to facilitate the efficient gathering of a CoE that may be admitted in court. The traditional DF processes are analogous in every digital environment, but in the IoT domain, they cause peculiarities throughout the capturing procedure. For this reason, the lifecycle for IoT-enabled investigation is split into a temporal phase (only if the evidence occurs in the IoT) as follows [43–49]: evidence identification, evidence collection, evidence preservation, evidence analysis, attack attribution, and evidence presentation.

3.1 Rationale Usage of IoT in Social Media Platforms

Because limited device memory occurs frequently, a degree of forensics is classified in the process of extracting evidence from IoT devices. Numerous devices, including various sensors, cameras for computer vision, RFIDs, network monitoring units, smart timers, and storage monitoring units, can be linked to this process in order to recognize and record the meta information of the node [50, 51]. Since several categories of devices are interconnected and give their functionality, obtaining or gathering a CoE—which is currently impractical—requires a high degree of device identification. Thus, network forensics is crucial to the entire process since it uses social media to trace device activity and extract certificate of origin (CoE) from network logs [51, 52]. Nevertheless, this integration distinguishes IoT-Network Forensics from the traditional one, supporting additional network models such as body area network (BAN)/personal area network (PAN) to maintain CoE. By effectively collaborating with the DF technique, we are able to schedule CoE for each and every social media channel account, including those with varying types of network protocols.

On the other hand, forensics examination is scheduled and imported based on a number of factors, including the identification and gathering of evidence as well as searches and seizures. Due to the smaller size and dimensional supportability of these devices in terms of memory preservation, IoT network forensics, which is used to identify the presence of IoT devices, is not always able to respond quickly in this situation [53], is therefore made to function both passively and independently. These specific stages play a crucial role in filtering and aggregating the multimedia data that has been collected, enabling it to be efficiently evaluated, reported, and presented in court for subsequent proceedings. Table 3 illustrates a systematic examination of the rationale usage of IoT in social media platforms [53–58], whose metrics of evaluation are mentioned as follows: (1) discussion of collaborative technological framework(s), (2) lack of privacy and security concerns, (3) lack of real-time investigation and incident response, (4) available opportunities (2012–2023), (5) SMF Modernized design and development requires, and (6) details argument based on requirement fulfillment (2013–2023).

Table 3. A systematic examination of rationale usage of IoT in social media platforms

Study	Discussion of collaborative technological framework(s)	Lack of privacy and security concerns	Lack of real-time investigation and incident response	Available opportunities (2013–2023)	SMF modernized design and developments requires	Detail argument based on requirement fulfillment
Mitra et al. [53]	D	P	P	√	Presented with partially discussion	√
Nowroozi et al. [54]	D	N	N	N/A	Presented with partially discussion	N/A
Mukta et al. [55]	P	D	D	√	Do not presented	√
Cover et al. [56]	N	D	P	√	Do not presented	√
Khoo et al. [57]	N	D	P	N/A	Presented with partially discussion	N/A
Zhang [58]	P	P	D	√	Do not presented	√

N = no discussion added, P = provide in-depth description, D = partially discussed.

3.2 Contextual Description of IoT in Deepfake Precaution

This work proposes a category to attempt a systematic review process in the context of IoT in deepfake detection and precaution. The goal is to gather all the empirical records that support a pre-specific eligibility technique to address a pre-defined research topic. These concerns are divided because they cover the pertinent review analysis prospects related to the IoT domain in deepfake [59–61]. This protocol outlines the state-of-the-arts that have included descriptions of deepfake and IoT technologies [53–60, 62–65]. The following questions are listed in order to answer the contextual scenario of this research to integrate IoT in deepfake detection and prevention:

Research Question 1: What is the exact research domain of the previously published state-of-the-arts?
Research Question 2: What the relevance we find to evaluate the particular research work?
Research Question 3: What are the major trends this research works derived?
Research Question 4: How the research changes over time?
Research Question 5: What parameters applied in such changes?

4. Cyber Security for Deepfake Detection

Since it is evident that deepfake technology is one of the tools most frequently employed for evil intent, a range of deepfake scams that are horrifying and disrupt the world through forgery attempts exist [66–71]. Here, we will highlight a few of these scams, which include the following [72–75]:

Election manipulation: A number of examples have been reported based on deepfake recordings that have gone viral after international leaders made comments to others. For example, Brack Obama's deepfakes have raised concerns about how this video would impact the US presidential election campaign.

Social engineering: Following election manipulation, social engineering schemes are another highly ranked deepfake category. In these scams, the audio is edited to trick listeners into thinking that reliable sources have stated any phrases that they did not.

Hoaxes and scams: Cybercriminals have recently exploited this technology to fabricate fraudulent claims, hoaxes, and phony items that weaken and destabilize organizations.

Identity theft: This refers to the use of technology in malicious attacks to create fraudulent documents, impersonate people, and create purchasing accounts in order to drive new identities and steal identities.

Automated misinformation attacks: In this way, technology is employed to provide false information about social issues and politics, as well as conspiracy theories.

Table 4 presents a systematic evaluation of application of cyber security for detection of deepfake [76–90], whose metrics of discussion are highlighted as follows: (1) normalized deepfakes, (2) induce cyber-attacks, (3) vishing, (4) business email compromise, (5) combating deepfakes, (6) role of blockchain DLT, (7) Involvement of AI/ML as the first line of deference, (8) third-party analyzer, (9) evaluator of deepfake generators, and (10) evaluator of deepfake discriminator.

Table 4. A systematic analysis of deepfake investigation and the role of social media forensics (2013–2023)

Context of discussion	**Year & study**										
	2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
	[76]	**[77]**	**[78]**	**[79, 80]**	**[81]**	**[82]**	**[83, 84]**	**[85]**	**[86, 87]**	**[88–90]**	**This work**
Normalized deepfakes	√	√	√		√	√	√		√	√	√
Induce cyber-attacks	√	√		√		√	√			√	√
Business email compromise	√	√	√	√	√				√	√	√
Combating deepfakes		√	√		√		√	√			√
Role of blockchain DLT		√	√	√					√		√
Involvement of AI/ML as the first line of deference			√	√		√	√		√		√
Evaluator of deepfake discriminator				√	√	√	√	√	√		√

4.1 The Role of Blockchain for Deepfake Prevention and Countermeasure in Cyberspace

To provide persistency, integrity, and transparency in the aforementioned deepfake detection process, the information of identified multimedia content and model-generated logs are recorded on the blockchain distributed ledger technology (BDLT) [91–95]. On the other hand, deepfakes are AI-enabled, phony multimedia representations that include motions like face swapping. The main idea is to use deepfake as a threat on social media to strengthen an environment of impunity where those who make deepfake face little to no repercussions for harming those who endure. Understanding the fundamentals of hashing, crypto/digital signatures, and consensus policies will help us resist this situation and take advantage of blockchain's potential to combat deepfake on social media channels. We can also adjust protection from and back to blockchain timestamping, which improves systems' real-time capabilities.

It is important to note that the benefits blockchain offers support in the areas of design and development. This entails converting current designs to the BDLT, which is unique in addressing various social media-related deepfake investigative architectures [96–99]. Blockchain timestamping does not require proof of ownership, but it can assist with a report's proof of stack. However, the BDLT has significant limitations that make it unable to detect deepfakes; in order to overcome these, AI-enabled

ML techniques must be integrated. It can aid in the offender's covert investigation and develop the capacity to recognize deepfake information on social media. The development of the platform, which combines BDLT and AI to detect deepfakes in a way that is effective, quick, and trustworthy, also calls for extensive cooperation between corporations, civil society, and the government technological community. Fig. 5 depicts the blockchain association's operational cycle.

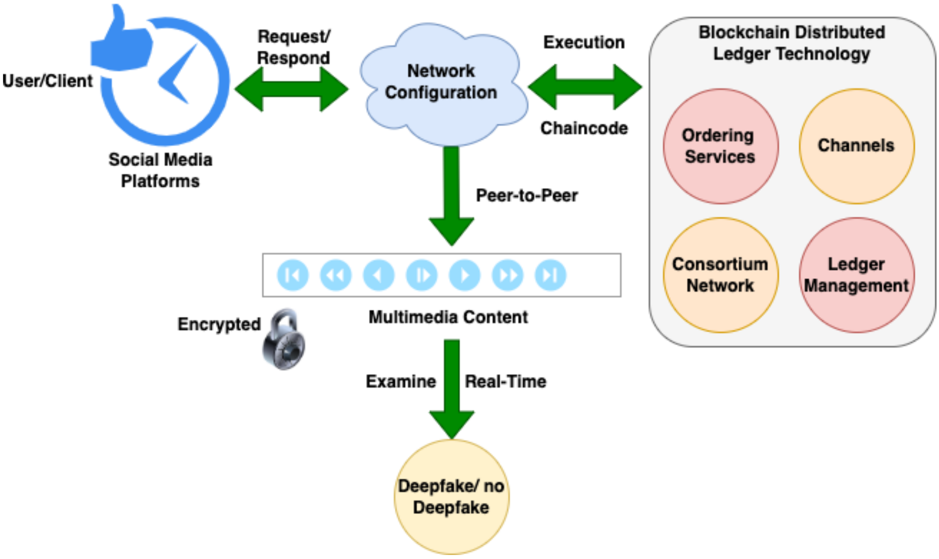

Fig. 5. Working cycle of blockchain for deepfake detection.

5. Digital Forensics in Multimedia Investigation

Digital forensic specialists are vital in the battle against multimedia deepfakes, particularly on social media sites [100]. Nonetheless, the following are some essential elements linked to the DF process that enhance real-time detection capability: research and development, dataset evaluation, algorithm analysis, authentication methods, and authentication procedures.

Research and development (R&D): Enhancing routine instruments, methods, and organizational structure for superior advancements.

Evaluation of the dataset: Deepfake models are analyzed and trained using various sources of gathered datasets.

Table 5. A systematic report to integrate digital forensics techniques in multimedia investigation and CoC hierarchy

Study	Discussion of collaborative technological framework(s)	Lack of standardization in the design of incident response lifecycle	Lack of real-time investigation and SMF	Available opportunities (2013–2023)	DF modernized design and developments requires	Detail argument based on requirement fulfillment (2013–2023)
Khan et al. [96]	D	N	P	√	Do not presented	N/A
Khan et al. [97]	P	N	P	N/A	Do not presented	N/A
Khan et al. [98]	N	P	P	N/A	Do not presented	√
Khan et al. [99]	N	D	U	N/A	Do not presented	√

N = no discussion added, P = provide in-depth description, D = partially discussed.

Authentication techniques: The validity, authenticity, and verification of multimedia materials must be achieved using pre-established ways.

Algorithm analysis: The analyzer helps evaluate deepfake models in order to comprehend their operations and activities.

Table 5 provides an overview of the methodical examination of technological integration, with a particular focus on the use of DF techniques in multimedia research and CoC management. The evaluation of these techniques is discussed below: (1) a conversation about cooperative technology framework(s), (2) inconsistency in the incidence response lifecycle design, (3) absence of real-time inquiry and SMF connectivity, (4) opportunities that are available (2013–2023), (5) needs for DF advancement design and development, and (6) detailed reasoning based on requirement fulfillment (2013–2023).

5.1 Survey, Search Section of Social Media Forensics, and Open Research Problems

In this section, we discuss some of the extensions that relate to the aforementioned prospects and how they can be integrated with DF's methods and tools to efficiently detect deepfakes in multimedia systems [99, 100]: audio analysis, facial artifact analysis, source data comparison, and metadata analysis. In the sections that follow, we address potential solutions for thwarting multimedia deepfakes on social media platforms [97, 98, 100–104].

Audio analysis: it is determined that there is a lack of consistency or delay in the deepfake sounds.

Analysis of face features artifacts: We discovered that deepfakes leave behind minute artifacts that are observable and worthy of follow-up research.

Comparison with preserved/collected data: In each case study, the recorded multimedia elements are contrasted with reference materials in order to identify any disparities.

Metadata analysis: Every time information is collected, it is examined for metadata that may indicate manipulation or forgery.

Table 6. Comparative analysis between classical, current, and modernized DF techniques (Test-1)

Metrics	DF technique (%)			Overall (%)
	Current	Classical	Modernized	
Manual investigation of deepfake	17.30	27.38	19.13	3.55
Incident response on deepfake	15.10	23.79	17.34	3.77
Real-time investigation of deepfake	13.80	21.59	15.78	2.21
Standardized lifecycle	11.70	17.93	11.96	1.90
Collaborative strategy (IoT, AI, blockchain, big data, and cyber security)	12.20	15.47	11.33	2.90

Table 7. Comparative analysis between classical, current, and modernized DF techniques (Test-2)

Metrics	DF technique (%)			Overall (%)
	Current	Classical	Modernized	
Manual investigation of deepfake	19.70	21.77	13.60	2.10
Incident response on deepfake	17.50	19.97	11.33	3.60
Real-time investigation of deepfake	15.70	17.31	9.11	3.30
Standardized lifecycle	13.81	13.11	7.13	1.93
Collaborative strategy (IoT, AI, blockchain, big data, and cyber security)	14.54	18.33	10.59	2.89

With respect to these mentioned aspects, Tables 6 and 7 [90–99] present the report of comparative analysis, which is scheduled between classical, current, and modernized DF techniques, whose evaluation metrics are highlighted as follows: (1) manual investigation of deepfake, (2) incident response on deepfake, (3) real-time investigation of deepfake, (4) standardized lifecycle, and (5) collaborative strategy (IoT, AI, blockchain, big data, and cyber security).

5.2 The Role of Blockchain for Deepfake Prevention and Countermeasure in Cyberspace

Because deepfakes are associated with black mirror, revenge porn, and fake news publication on social media, they are not considered acceptable evidence to be used in court [99, 100]. There is a worry over how a lawyer can keep up with a case summary with such staff, who can assess which sections pertain to the accused, and how the legal system might handle this. These queries bring up unresolved research issues for technological experts. Furthermore, the most important thing is to figure out how to handle this. The best way to respond to this incident is to store identified content securely so that it can be shown in court, like CoC. Using basic picture editing software like Photoshop, this approach identifies and categorizes altered or edited contents, swapped or faked face photos, and body doubling. It then stores each of them in a different chunk. This is because it does not require expensive investigation or analysis techniques; Illustrator or Photoshop's retouching feature can be used to remove it. Investigating bogus information that is spreading through social media channels, such as election manipulation and deception, social engineering, hoaxes and frauds, identity theft, automated disinformation attacks, etc., is the main motivation behind designing CoC for deepfake, as Fig. 6 illustrates. More investigative units are needed for this kind of activity in order to record it in a protected, safe, and secure manner. The following elements are highlighted in the list of factors to assess deepfake in multimedia systems (as shown in Fig. 6): (1) original image/video capture, (2) event collection from various external legitimate sources, (3) database maintenance, (4) application of AI-enabled DL/ML approaches (e.g., ANN), feature extraction and classification, (5) identification of fake/real material, and (6) deepfake detection. Nonetheless, the following is indicated regarding the established procedure for CoC administration and optimization: identification, capturing, examination (filtering), analysis (aggregate), preservation, and documentation.

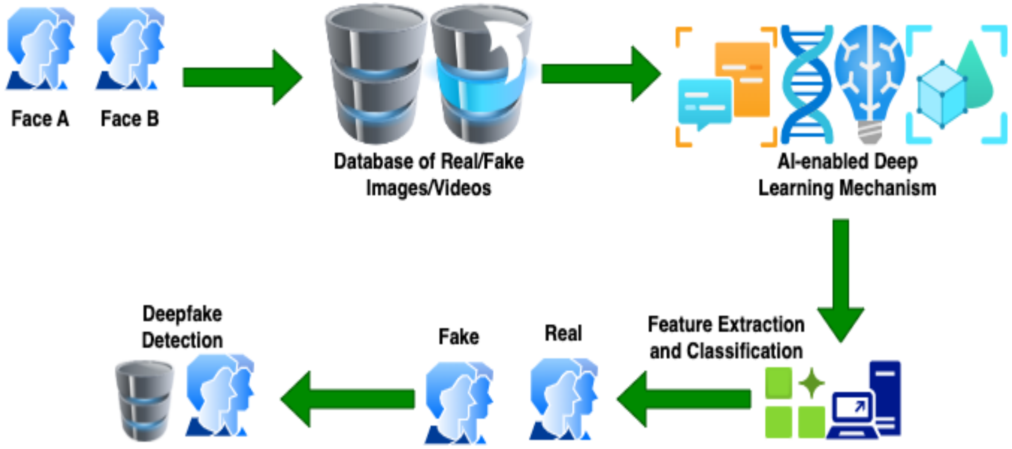
Face A
Face B
Database of Real/Fake Images/Videos
AI-enabled Deep Learning Mechanism
Feature Extraction and Classification
Real
Fake
Deepfake Detection

Fig. 6. Process of deepfake investigation.

6. Conclusion

This review study primarily focuses on the gaps in technology development as well as the intricacies of deepfake detection and its prevalence in current social media platforms. To that end, we conduct a thorough analysis of more than 130 credible, state-of-the-art research articles in order to pinpoint unmet research needs related to the impacts of deepfakes and recommended countermeasures. These suggestions are meant to lessen the limitations of social media networks. This is a result of the SMF being presented with a significant advancement in digital investigative techniques, incident response, and regulatory compliance. Additionally, this review article evaluates numerous state-of-the-art frameworks where experts describe different approaches to detecting deepfakes on social media platforms but are restricted to the conventional hierarchy of analysis. Consequently, the study separates the research that aims to combat deepfake in real-time and determines the technologies required to satisfy those demands. To make SMF-enabled deepfake investigation more dependable, quick, efficient, and successful, BDLT, IoT, cyber security, and AI are successfully integrated. Furthermore, the cooperation of new strategies and traditional algorithms to extract more complicated opportunities while transforming the current drawbacks mentioned in the architectural upgrades. In order to foster technological developments and maturity, this paper also identifies a few open research topics and investigates viable solutions. It is also partitioned so that technology experts can devote more of their attention to making it more dependable and effective. However, in the upcoming years, a number of deepfake experts plan to provide their proposals detailing the current research gaps and the fulfillment of architectural requirements. They benefit from this study in this way. As a result, we state that this work is a great fit for those researching multimedia forensics to build a new, secure architecture for further studies and advancements on real-time deepfake research.

Author's Contributions

Conceptualization, AAK, AAL, RB; Investigation and methodology, RA, SA; Project administration, AAK, JAJA, AAL; Resources, RA, SA, AMB; Supervision, AAK, JAJA; Writing of the original draft, AAK; Writing of the review and editing, JAJA; Software, RA, SA, AMB; Validation, RA, SA, AMB; Formal analysis, RA, SA, AMB; Data curation, AAK, JAJA, AAL; Visualization, RB, JAJA, AAK, AAL.

Funding

This research was supported by the Universiti Teknikal Malaysia Melaka (UTeM) and the Ministry of Higher Education of Malaysia (MOHE); and the Taif University, Saudi Arabia (Project No. TU-DSPP-2024-17).

Competing Interests

The authors declare that they have no competing interests.

References

[1] M. Sharma and M. Kaur, "A review of Deepfake technology: an emerging AI threat," in *Soft Computing for Security Applications*. Singapore: Springer, 2022, pp. 605-619. https://doi.org/10.1007/978-981-16-5301-8_44

[2] D. Mao, S. Zhao, and Z. Hao, "A shared updatable method of content regulation for deepfake videos based on blockchain," *Applied Intelligence*, vol. 52, no. 13, pp. 15557-15574, 2022. https://doi.org/10.1007/s10489-021-03156-x

[3] D. Nagothu, R. Xu, Y. Chen, E. Blasch, and A. Aved, "DeFakePro: decentralized deepfake attacks detection using ENF authentication," *IT Professional*, vol. 24, no. 5, pp. 46-52, 2022. https://doi.org/10.1109/MITP.2022.3172653

[4] M. Taeb, H. Chi, and S. Bernadin, "Digital evidence acquisition and deepfake detection with decentralized applications," in *Proceedings of PEARC: Practice and Experience in Advanced Research Computing*, 2022, pp. 1-2. https://doi.org/10.1145/3491418.3535127

[5] L. Zhang, T. Qiao, M. Xu, N. Zheng, and S. Xie, “Unsupervised learning-based framework for deepfake video detection,” *IEEE Transactions on Multimedia*, vol. 25, pp. 4785-4799, 2022. https://doi.org/10.1109/TMM.2022.3182509

[6] H. Asgharzadeh, A. Ghaffari, M. Masdari, and F. S. Gharehchopogh, “Anomaly-based intrusion detection system in the Internet of Things using a convolutional neural network and multi-objective enhanced capuchin search algorithm,” *Journal of Parallel and Distributed Computing*, vol. 175, pp. 1-21, 2023. https://doi.org/10.1016/j.jpdc.2022.12.009

[7] A. A. Khan, A. A. Laghari, Z. A. Shaikh, Z. Dacko-Pikiewicz, and S. Kot, “Internet of Things (IoT) security with blockchain technology: a state-of-the-art review,” *IEEE Access*, vol. 10, pp. 122679-122695, 2022. https://doi.org/10.1109/ACCESS.2022.3223370

[8] F. S. Gharehchopogh, B. Abdollahzadeh, S. Barshandeh, and B. Arasteh, “A multi-objective mutation-based dynamic Harris Hawks optimization for botnet detection in IoT,” *Internet of Things*, vol. 24, article no. 100952, 2023. https://doi.org/10.1016/j.iot.2023.100952

[9] K. Moghaddasi, S. Rajabi, and F. S. Gharehchopogh, “Multi-objective secure task offloading strategy for blockchain-enabled IoV-MEC systems: a double deep Q-network approach,” *IEEE Access*, vol. 12, pp. 3437-3463, 2024. https://doi.org/10.1109/ACCESS.2023.3348513

[10] A. A. Khan, A. A. Laghari, A. A. Shaikh, M. A. Dootio, V. V. Estrela, and R. T. Lopes, “A blockchain security module for brain-computer interface (BCI) with multimedia life cycle framework (MLCF),” *Neuroscience Informatics*, vol. 2, no. 1, article no. 100030, 2022. https://doi.org/10.1016/j.neuri.2021.100030

[11] F. Hosseini, F. S. Gharehchopogh, and M. Masdari, “A botnet detection in IoT using a hybrid multi-objective optimization algorithm,” *New Generation Computing*, vol. 40, no. 3, pp. 809-843, 2022. https://doi.org/10.1007/s00354-022-00188-w

[12] M. Gaikwad and A. Dhutonde, “Statistical analysis of deep learning models used for social media forensics from an empirical perspective,” *International Journal of Scientific Research in Science and Technology*, vol. 10, no. 5, pp. 46-64, 2023.

[13] M. Tampubolon, “Digital face forgery and the role of digital forensics,” *International Journal for the Semiotics of Law*, vol. 37, no. 3, pp. 753-767, 2024. https://doi.org/10.1007/s11196-023-10030-1

[14] A. A. Khan, A. A. Laghari, and S. A. Awan, “Machine learning in computer vision: a review,” *EAI Endorsed Transactions on Scalable Information Systems*, vol. 8, no. 32, article no. e4, 2021. https://doi.org/10.4108/eai.21-4-2021.169418

[15] A. Purwadi, C. Y. Serfiyani, and C. R. Serfiyani, “Legal landscape on national cybersecurity capacity in combating cyberterrorism using deep fake technology in Indonesia,” *International Journal of Cyber Criminology*, vol. 16, no. 1, pp. 123-140, 2022.

[16] A. Aldweesh, “The impact of blockchain on digital content distribution: a systematic review,” *Wireless Networks*, vol. 30, no. 2, pp. 763-779, 2024. https://doi.org/10.1007/s11276-023-03524-0

[17] A. A. Khan, S. Bourouis, M. M. Kamruzzaman, M. Hadjouni, Z. A. Shaikh, A. A. Laghari, H. Elmannai, and S. Dhahbi, “Data security in healthcare industrial internet of things with blockchain,” *IEEE Sensors Journal*, vol. 23, no. 20, pp. 25144-25151, 2023. https://doi.org/10.1109/JSEN.2023.3273851

[18] K. Narayan, H. Agarwal, S. Mittal, K. Thakral, S. Kundu, M. Vatsa, and R. Singh, “DeSI: deepfake source identifier for social media,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 2857-2866. https://doi.org/10.1109/CVPRW56347.2022.00323

[19] N. O'Donnell, “Have we no decency? Section 230 and the liability of social media companies for deepfake videos,” 2021 [Online]. Available: https://illinoislawreview.org/print/vol-2021-no-2/have-we-no-decency/.

[20] A. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, “A machine learning based approach for deepfake detection in social media through key video frame extraction,” *SN Computer Science*, vol. 2, no. 2, article no. 98, 2021. https://doi.org/10.1007/s42979-021-00495-x

[21] S. M. Saravani, I. Ray, and I. Ray, “Automated identification of social media bots using deepfake text detection,” in *Information Systems Security*. Cham, Switzerland: Springer, 2021, pp. 111-123. https://doi.org/10.1007/978-3-030-92571-0_7

[22] A. Chadha, V. Kumar, S. Kashyap, and M. Gupta, “Deepfake: an overview,” in *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*, Singapore: Springer, 2021, pp. 557-566. https://doi.org/10.1007/978-981-16-0733-2_39

[23] M. Kumar and H. K. Sharma, "A GAN-based model of deepfake detection in social media," *Procedia Computer Science*, vol. 218, pp. 2153-2162, 2023. https://doi.org/10.1016/j.procs.2023.01.191

[24] A. O. Kwok and S. G. Koh, "Deepfake: a social construction of technology perspective," *Current Issues in Tourism*, vol. 24, no. 13, pp. 1798-1802, 2021. https://doi.org/10.1080/13683500.2020.1738357

[25] T. Fagni, F. Falchi, M. Gambini, A. Martella, and M. Tesconi, "TweepFake: about detecting deepfake tweets," *PLOS One*, vol. 16, no. 5, article no. e0251415, 2021. https://doi.org/10.1371/journal.pone.0251415

[26] A. Yadlin-Segal and Y. Oppenheim, "dystopia is it anyway? Deepfakes and social media regulation," *Convergence*, vol. 27, no. 1, pp. 36-51, 2021. https://doi.org/10.1177/1354856520923963

[27] S. Ahmed, "Navigating the maze: deepfakes, cognitive ability, and social media news skepticism," *New Media & Society*, vol. 25, no. 5, pp. 1108-1129, 2023. https://doi.org/10.1177/14614448211019198

[28] J. T. Hancock and J. N. Bailenson, "The social impact of deepfakes," *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 3, pp. 149-152, 2021. https://doi.org/10.1089/cyber.2021.29208.jth

[29] A. Kyrpa, "Social media as a tool for the formation of media literacy," *Viae Educationis*, vol. 1, no. 4, pp. 22-29, 2013.

[30] A. Ianchenko, "Mediatisation and self-mediation of political participation: new citizenship practices and social media," *Bulletin of Mariupol State University Series History Political Science*, vol. 4, no. 10, pp. 133-141, 2014.

[31] B. Moriarty, "Defeating ISIS on Twitter," *Technology Science*, 2015 [Online]. Available: https://techscience.org/a/2015092904/.

[32] L. M. Jones and K. J. Mitchell, "Defining and measuring youth digital citizenship," *New Media & Society*, vol. 18, no. 9, pp. 2063-2079, 2016. https://doi.org/10.1177/1461444815577797

[33] P. Arora and L. Scheiber, "Slumdog romance: Facebook love and digital privacy at the margins," *Media, Culture & Society*, vol. 39, no. 3, pp. 408-422, 2017. https://doi.org/10.1177/0163443717691225

[34] J. Fletcher, "Deepfakes, artificial intelligence, and some kind of dystopia: the new faces of online post-fact performance," *Theatre Journal*, vol. 70, no. 4, pp. 455-471, 2018. https://doi.org/10.1353/tj.2018.0097

[35] M. Westerlund, "The emergence of deepfake technology: a review," *Technology Innovation Management Review*, vol. 9, no. 11, pp. 40-53, 2019. http://doi.org/10.22215/timreview/1282

[36] E. Perot and F. Mostert, "Fake it till you make it: an examination of the US and English approaches to persona protection as applied to deepfakes on social media," *Journal of Intellectual Property Law & Practice*, vol. 15, no. 1, pp. 32-39, 2020. https://doi.org/10.1093/jiplp/jpz164

[37] B. F. Nasar, T. Sajini, and E. R. Lason, "Deepfake detection in media files-audios, images and videos," in *Proceedings of 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Thiruvananthapuram, India, 2020, pp. 74-79. https://doi.org/10.1109/RAICS51191.2020.9332516

[38] P. Neekhara, B. Dolhansky, J. Bitton, and C. C. Ferrer, "Adversarial threats to deepfake detection: a practical perspective," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 2021, pp. 923-932. https://doi.org/10.1109/CVPRW53098.2021.00103

[39] Y. Lee, K. T. Huang, R. Blom, R. Schriner, and C. A. Ciccarelli, "To believe or not to believe: framing analysis of content and audience response of top 10 deepfake videos on Youtube," *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 3, pp. 153-158, 2021. https://doi.org/10.1089/cyber.2020.0176

[40] S. Ahmed, "Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes," *Personality and Individual Differences*, vol. 182, article no. 111074, 2021. https://doi.org/10.1016/j.paid.2021.111074

[41] G. Murphy and E. Flynn, "Deepfake false memories," *Memory*, vol. 30, no. 4, pp. 480-492, 2022. https://doi.org/10.1080/09658211.2021.1919715

[42] M. Groh, Z. Epstein, C. Firestone, and R. Picard, "Deepfake detection by human crowds, machines, and machine-informed crowds," *Proceedings of the National Academy of Sciences*, vol. 119, no. 1, article no. e2110013119, 2022. https://doi.org/10.1073/pnas.2110013119

[43] A. A. Khan, A. A. Wagan, A. A. Laghari, A. R. Gilal, I. A. Aziz, and B. A. Talpur, "BIoMT: a state-of-the-art consortium serverless network architecture for healthcare system using blockchain smart contracts," *IEEE Access*, vol. 10, pp. 78887-78898, 2022. https://doi.org/10.1109/ACCESS.2022.3194195

[44] S. Y. Shin and J. Lee, "The effect of deepfake video on news credibility and corrective influence of cost-based knowledge about deepfakes," *Digital Journalism*, vol. 10, no. 3, pp. 412-432, 2022. https://doi.org/10.1080/21670811.2022.2026797

[45] A. A. Khan, A. A., Laghari, M. Shafiq, O. Cheikhrouhou, W. Alhakami, H. Hamam, and Z. A. Shaikh, "Healthcare ledger management: a blockchain and machine learning-enabled novel and secure architecture for medical industry," *Human-centric Computing and Information Sciences*, vol. 12, article no. 55, 2022. http://dx.doi.org/10.22967/HCIS.2022.12.055

[46] S. D. Bray, S. D. Johnson, and B. Kleinberg, "Testing human ability to detect 'deepfake' images of human faces," *Journal of Cybersecurity*, vol. 9, no. 1, pp. 1-18, 2023. https://doi.org/10.1093/cybsec/tyad011

[47] A. A. Khan, A. A. Laghari, M. Rashid, H. Li, A. R. Javed, and T. R. Gadekallu, "Artificial intelligence and blockchain technology for secure smart grid and power distribution automation: a state-of-the-art review," *Sustainable Energy Technologies and Assessments*, vol. 57, article no. 103282, 2023. https://doi.org/10.1016/j.seta.2023.103282

[48] Y. Patel, S. Tanwar, R. Gupta, P. Bhattacharya, I. E. Davidson, R. Nyameko, S. Aluvala, and V. Vimal, "Deepfake generation and detection: case study and challenges," *IEEE Access*, vol. 11, pp. 143296-143323, 2023. https://doi.org/10.1109/ACCESS.2023.3342107

[49] Z. Xia, T. Qiao, M. Xu, X. Wu, L. Han, and Y. Chen, "Deepfake video detection based on MesoNet with preprocessing module," *Symmetry*, vol. 14, no. 5, article no. 939, 2022. https://doi.org/10.3390/sym14050939

[50] Y. L. Ng, "An error management approach to perceived fakeness of deepfakes: the moderating role of perceived deepfake targeted politicians' personality characteristics," *Current Psychology*, vol. 42, no. 29, pp. 25658-25669, 2023. https://doi.org/10.1007/s12144-022-03621-x

[51] L. Whittaker, R. Mulcahy, K. Letheren, J. Kietzmann, and R. Russell-Bennett, "Mapping the deepfake landscape for innovation: a multidisciplinary systematic review and future research agenda," *Technovation*, vol. 125, article no. 102784, 2023. https://doi.org/10.1016/j.technovation.2023.102784

[52] A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, "DeepFake detection for human face images and videos: a survey," *IEEE Access*, vol. 10, pp. 18757-18775, 2022. https://doi.org/10.1109/ACCESS.2022.3151186

[53] A. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, "EasyDeep: an IoT friendly robust detection method for GAN generated deepfake images in social media," in *Internet of Things: Technology and Application*. Cham, Switzerland: Springer, 2021, pp. 217-236. https://doi.org/10.1007/978-3-030-96466-5_14

[54] E. Nowroozi, S. Seyedshoari, M. Mohammadi, and A. Jolfaei, "Impact of media forensics and deepfake in society," in *Breakthroughs in Digital Biometrics and Forensics*. Cham, Switzerland: Springer, 2022, pp. 387-410. https://doi.org/10.1007/978-3-031-10706-1_18

[55] M. S. H. Mukta, J. Ahmad, M. A. K. Raiaan, S. Islam, S. Azam, M. E. Ali, and M. Jonkman, "An investigation of the effectiveness of deepfake models and tools," *Journal of Sensor and Actuator Networks*, vol. 12, no. 4, article no. 61, 2023. https://doi.org/10.3390/jsan12040061

[56] R. Cover, "Deepfake culture: the emergence of audio-video deception as an object of social anxiety and regulation," *Continuum*, vol. 36, no. 4, pp. 609-621, 2022. https://doi.org/10.1080/10304312.2022.2084039

[57] B. Khoo, R. C. W. Phan, and C. H. Lim, "Deepfake attribution: on the source identification of artificially generated images," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 3, article no. e1438, 2022. https://doi.org/10.1002/widm.1438

[58] T. Zhang, "Deepfake generation and detection, a survey," *Multimedia Tools and Applications*, vol. 81, no. 5, pp. 6259-6276, 2022. https://doi.org/10.1007/s11042-021-11733-y

[59] N. Kshetri, "The economics of deepfakes," *Computer*, vol. 56, no. 8, pp. 89-94, 2023. https://doi.org/10.1109/MC.2023.3276068

[60] H. Lu and H. Chu, "Let the dead talk: how deepfake resurrection narratives influence audience respons e in prosocial contexts," *Computers in Human Behavior*, vol. 145, article no. 107761, 2023. https://doi.org/10.1016/j.chb.2023.107761

[61] J. Lee and S. Y. Shin, "Something that they never said: multimodal disinformation and source vividness in understanding the power of AI-enabled deepfake news," *Media Psychology*, vol. 25, no. 4, pp. 531-546, 2022. https://doi.org/10.1080/15213269.2021.2007489

[62] X. Wang, J. Huang, S. Ma, S. Nepal, and C. Xu, "Deepfake disrupter: the detector of deepfake is my friend," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 14900-14909. https://doi.org/10.1109/CVPR52688.2022.01450

[63] F. S. Sturino, "Deepfake technology and individual rights," *Social Theory and Practice*, vol. 49, no. 1, pp. 161-187, 2023. https://doi.org/10.5840/soctheorpract2023310184

[64] X. Liao, Y. Wang, T. Wang, J. Hu, and X. Wu,": facial muscle motions for detecting compressed deepfake videos over social networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 12, pp. 7236-7251, 2023. https://doi.org/10.1109/TCSVT.2023.3278310

[65] I. Temnikova, I. Marinova, S. Gargova, R. Margova, and I. Koychev, "Looking for traces of textual deepfakes in Bulgarian on social media," in *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing (RANLP)*, Varna, Bulgaria, 2023, pp. 1151-1161.

[66] P. Korshunov and S. Marcel, "Improving generalization of deepfake detection with data farming and few-shot learning," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 3, pp. 386-397, 2022. https://doi.org/10.1109/TBIOM.2022.3143404

[67] S. Hussain, P. Neekhara, B. Dolhansky, J. Bitton, C. C. Ferrer, J. McAuley, and F. Koushanfar, "Exposing vulnerabilities of deepfake detection systems with robust attacks," *Digital Threats: Research and Practice (DTRAP)*, vol. 3, no. 3, article no. 30, 2022. https://doi.org/10.1145/3464307

[68] B. Dash and P. Sharma, "Are ChatGPT and deepfake algorithms endangering the cybersecurity industry? A review," *International Journal of Engineering and Applied Sciences*, vol. 10, no. 1, pp. 21-39, 2023.

[69] Y. Ju, S. Hu, S. Jia, G. H. Chen, and S. Lyu, "Improving fairness in deepfake detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, 2024, pp. 4643-4653. https://doi.org/10.1109/WACV57701.2024.00459

[70] I. Sharma, K. Jain, A. Behl, A. Baabdullah, M. Giannakis, and Y. Dwivedi, "Examining the motivations of sharing political deepfake videos: the role of political brand hate and moral consciousness," *Internet Research*, vol. 33, no. 5, pp. 1727-1749, 2023. https://doi.org/10.1108/INTR-07-2022-0563

[71] D. Fido, J. Rao, and C. A. Harper, "Celebrity status, sex, and variation in psychopathy predicts judgements of and proclivity to generate and distribute deepfake pornography," *Computers in Human Behavior*, vol. 129, article no. 107141, 2022. https://doi.org/10.1016/j.chb.2021.107141

[72] L. Guarnera, O. Giudice, F. Guarnera, A. Ortis, G. Puglisi, A. Paratore, et al., "The face deepfake detection challenge," *Journal of Imaging*, vol. 8, no. 10, article no. 263, 2022. https://doi.org/10.3390/jimaging8100263

[73] A. Aghasanli, D. Kangin, and P. Angelov, "Interpretable-through-prototypes deepfake detection for diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, France, 2023, pp. 467-474. https://doi.org/10.1109/ICCVW60793.2023.00053

[74] D. Siegel, C. Kratzer, S. Seidlitz, and J. Dittmann, "Forensic data model for artificial intelligence based media forensics-Illustrated on the example of DeepFake detection," *Electronic Imaging*, vol. 34, article no. MWSF-324, 2022. https://doi.org/10.2352/EI.2022.34.4.MWSF-324

[75] T. Wang and K. P. Chow, "Noise based deepfake detection via multi-head relative-interaction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 12, pp. 14548-14556, 2023. https://doi.org/10.1609/aaai.v37i12.26701

[76] J. Van Dijck, "Facebook and the engineering of connectivity: a multi-layered approach to social media platforms," *Convergence*, vol. 19, no. 2, pp. 141-155, 2013. https://doi.org/10.1177/1354856512457548

[77] M. Button, C. M. Nicholls, J. Kerr, and R. Owen, "Online frauds: learning from victims why they fall for these scams," *Australian & New Zealand Journal of Criminology*, vol. 47, no. 3, pp. 391-408, 2014. https://doi.org/10.1177/0004865814521224

[78] H. Kennedy and G. Moss, "Known or knowing publics? Social media data mining and the question of public agency," *Big Data & Society*, vol. 2, no. 2, article no. 2053951715611145, 2015. https://doi.org/10.1177/2053951715611145

[79] A. Fotopoulou, "Digital and networked by default? Women's organisations and the social imaginary of networked feminism," *New Media & Society*, vol. 18, no. 6, pp. 989-1005, 2016. https://doi.org/10.1177/1461444814552264

[80] A. Dodge, "Digitizing rape culture: Online sexual violence and the power of the digital photograph," *Crime, Media, Culture*, vol. 12, no. 1, pp. 65-82, 2016. https://doi.org/10.1177/1741659015601173

[81] K. Albury, J. Burgess, B. Light, K. Race, and R. Wilken, "Data cultures of mobile dating and hook-up apps: emerging issues for critical social science research," *Big Data & Society*, vol. 4, no. 2, article no. 2053951717720950, 2017. https://doi.org/10.1177/2053951717720950

[82] D. Guera and E. J. Delp, “Deepfake video detection using recurrent neural networks,” in *Proceedings of 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Auckland, New Zealand, 2018, pp. 1-6. https://doi.org/10.1109/AVSS.2018.8639163

[83] A. A. Khan, J. Yang, A. A. Laghari, A. M. Baqasah, R. Alroobaea, C. S. Ku, R. Alizadehsani, U. R. Acharya, and L. Y. Por, “BAIoT-EMS: consortium network for small-medium enterprises management system with blockchain and augmented intelligence of things,” *Engineering Applications of Artificial Intelligence*, vol. 141, article no. 109838, 2025. https://doi.org/10.1016/j.engappai.2024.109838

[84] B. Chesney and D. Citron, “Deep fakes: a looming challenge for privacy, democracy, and national security,” *California Law Review*, vol. 107, no. 6, pp. 1753-1820, 2019. https://doi.org/10.15779/Z38RV0D15J

[85] C. P. F. Repez and M. M. Popescu, “Social media and the threats against human security deepfake and fake news,” in *Romanian Military Thinking*. Bucuresti, Romania: Centrul tehnic-editorial al armatei, 2020, pp. 44-55.

[86] H. Chi, U. Maduakor, R. Alo, and E. Williams, “Integrating deepfake detection into cybersecurity curriculum,” in *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 1*. Cham, Switzerland: Springer, 2021, pp. 588-598. https://doi.org/10.1007/978-3-030-63128-4_45

[87] B. C. Taylor, “Defending the state from digital Deceit: the reflexive securitization of deepfake,” *Critical Studies in Media Communication*, vol. 38, no. 1, pp. 1-17, 2021. https://doi.org/10.1080/15295036.2020.1833058

[88] A. Raza, K. Munir, and M. Almutairi, “A novel deep learning approach for deepfake image detection,” *Applied Sciences*, vol. 12, no. 19, article no. 9820, 2022. https://doi.org/10.3390/app12199820

[89] V. V. V. N. S. Vamsi, S. S. Shet, S. S M. Reddy, S. S. Rose, S. R. Shetty, S. Sathvika, M. S. Supriya, and S. P. Shankar, “Deepfake detection in digital media forensics,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 74-79, 2022. https://doi.org/10.1016/j.gltp.2022.04.017

[90] M. Taeb and H. Chi, “Comparison of deepfake detection techniques through deep learning,” *Journal of Cybersecurity and Privacy*, vol. 2, no. 1, pp. 89-106, 2022. https://doi.org/10.3390/jcp2010007

[91] C. Nastasi, “Multimedia forensics: from image manipulation to the deep fake: new threats in the social media era,” Ph.D. dissertation, University of Catania, Catania, Italy, 2021.

[92] M. Khari, A. K. Garg, A. H. Gandomi, R. Gupta, R. Patan, and B. Balusamy, “Securing data in Internet of Things (IoT) using cryptography and steganography technique,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 1, pp. 73-80, 2020. https://doi.org/10.1109/TSMC.2019.2903785

[93] A. A. Khan, A. A. Laghari, A. M. Baqasah, R. Bacarra, R. Alroobaea, M. Alsafyani, and J. A. J. Alsayaydeh, “BDLT-IoMT: a novel architecture: SVM machine learning for robust and secure data processing in Internet of Medical Things with blockchain cybersecurity,” *The Journal of Supercomputing*, vol. 81, no. 1, article no. 271, 2025. https://doi.org/10.1007/s11227-024-06782-7

[94] J. A. J. Alsayaydeh, W. A. Indra, A. W. Y. Khang, V. Shkarupylo, and D. A. P. P. Jkatisan, “Development of vehicle ignition using fingerprint,” *ARPN Journal of Engineering and Applied Sciences*, vol. 14, no. 23, pp. 4045-4053, 2019.

[95] S. Nigam, U. Sugandh, and M. Khari, “The integration of blockchain and IoT edge devices for smart agriculture: challenges and use cases,” *Advances in computers*, vol. 127, pp. 507-537, 2022. https://doi.org/10.1016/bs.adcom.2022.02.015

[96] N. A. Afifie, A. W. Y. Khang, A. S. B. Ja'afar, A. F. B. M. Amin, J. A. J. Alsayaydeh, W. A. Indra, S. G. Herawan, and A. B. Ramli, “Evaluation method of mesh protocol over ESP32 and ESP8266,” *Baghdad Science Journal*, vol. 18, no. 4, pp. 1398-1401, 2021. https://doi.org/10.21123/bsj.2021.18.4(Suppl.).1397

[97] S. Ferreira, M. Antunes, and M. E. Correia, “Exposing manipulated photos and videos in digital forensics analysis,” *Journal of Imaging*, vol. 7, no. 7, article no. 102, 2021. https://doi.org/10.3390/jimaging7070102

[98] A. A. Khan, A. A. Laghari, H. Elmannai, A. A. Shaikh, S. Bourouis, M. Hadjouni, and R.Alroobaea, “GAN-IoTVS: a novel internet of multimedia things-enabled video streaming compression model using GAN and fuzzy logic,” *IEEE Sensors Journal*, vol. 23, no. 23, pp. 29434-29441, 2023. https://doi.org/10.1109/JSEN.2023.3316088

[99] A. A. Khan, A. A. Shaikh, Z. A. Shaikh, A. A. Laghari, and S. Karim, “IPM-Model: AI and metaheuristic-enabled face recognition using image partial matching for multimedia forensics investigation with genetic algorithm,” *Multimedia Tools and Applications*, vol. 81, no. 17, pp. 23533-23549, 2022. https://doi.org/10.1007/s11042-022-12398-x

[100] J. A. J. Alsayayadeh, M. F. Yusof, M. Z. Abdul Halim, M. N. S. Zainudin and S. G. Herawan, "Patient health monitoring system development using ESP8266 and Arduino with IoT platform" *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 4, pp. 617-624, 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140467

[101] A. A. Khan, A. A. Shaikh, and A. A. Laghari, "IoT with multimedia investigation: a secure process of digital forensics chain-of-custody using blockchain Hyperledger sawtooth," *Arabian Journal for Science and Engineering*, vol. 48, no. 8, pp. 10173-10188, 2023. https://doi.org/10.1007/s13369-022-07555-1

[102] A. A. Khan, X. Zhang, F. Hajjej, J. Yang, C. S. Ku, and L. Y. Por, "ASMF: ambient social media forensics chain of custody with an intelligent digital investigation process using federated learning," *Heliyon*, vol. 10, no. 1, article no. e23254, 2024. https://doi.org/10.1016/j.heliyon.2023.e23254

[103] S. Anitha, N. Anitha, N. Ashok, T. Daranya, B. Nandhini, and V. Chandrasekaran, "Detection of deepfakes in financial transactions using algorand blockchain consensus mechanism," in *Proceedings of International Conference on Network Security and Blockchain Technology*. Singapore: Springer, 2023, pp. 173-183. https://doi.org/10.1007/978-981-99-4433-0_15

[104] V. Shkarupylo, I. Blinov, A. Chemeris, V. Dusheba, J. A. J. Alsayaydeh and A. Oliinyk, "Iterative approach to TLC model checker application," in *Proceedings of 2021 IEEE 2nd KhPI Week on Advanced Technology (KhPIWeek)*, Kharkiv, Ukraine, 2021, pp. 283-287. https://doi.org/10.1109/KhPIWeek53812.2021.9570055