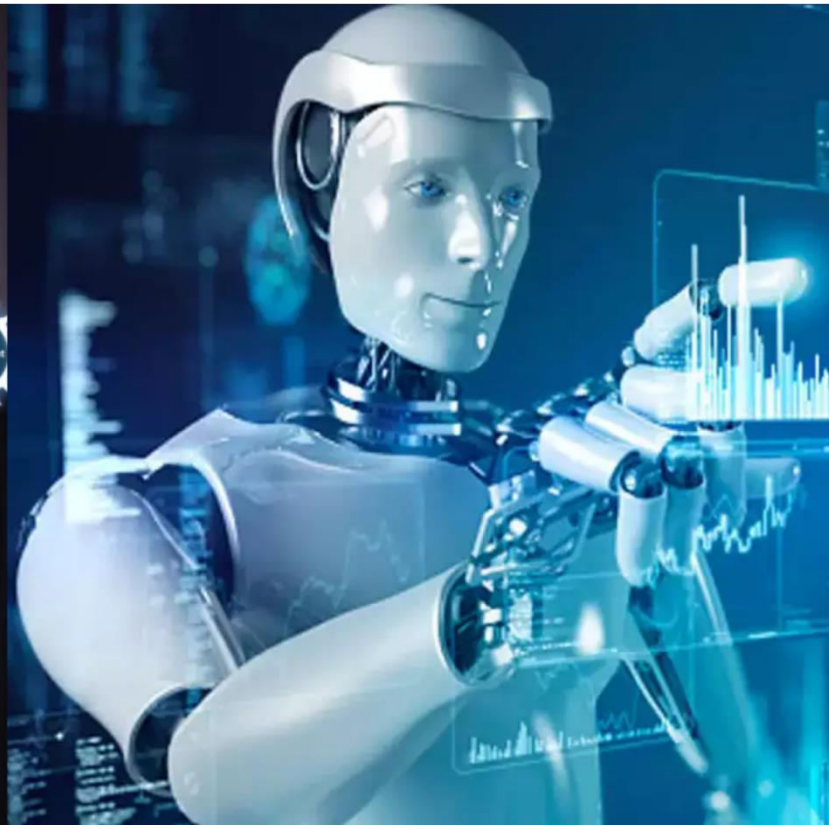
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 2, February 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402003|

Phishing 2.0: Leveraging Cyber Threat Intelligence to Combat Deepfake-Enhanced Social Engineering

Alex Mathew [1], Rodas, Federico[2]

Department of Cybersecurity, Bethany College, West Virginia, USA[1,2]

ABSTRACT: Standard phishing techniques have continued to improve in sophistication and devastating effect, enhanced by Deepfake Technology, referred to as Phishing 2.0. Due to enhanced realistic video and voice alterations, deepfakes have greatly improved the prospects of social engineering attacks so that they become challenging to detect. All these prolonged strategies present grave difficulties to conventional security models, proving why sophisticated Cyber Threat Intelligence (CTI) systems are necessary. Rising cyber technologies, AI-supported tools, and Deepfake technologies conclude that Cybersecurity Policies must be adapted. AI, in cybersecurity, plays a crucial role in these advanced threats. By progressing with the CTI frameworks even more, organizations can leverage AI and Machine Learning (ML) capabilities to predict, identify threats, and anticipate potential risks.

KEYWORDS: Phishing 2.0, Deepfake technology, Social Engineering, Cyber Threat Intelligence(CTI), Cybersecurity Policies, Deepfake Detection, AI tools, CTI frameworks and Machine Learning.

I. INTRODUCTION

Phishing is a form of identity theft that, in the past, used emails and clones of legitimate, genuine websites to con people into revealing sensitive information (Alabdan, 2020). Phishing, a highly used malicious attack, has developed over the years. Recently, their modifications have been served with the help of deepfake techniques. The so-called deepfakes are AI-generated synthetic videos or audio impersonating a real person. Such innovations have brought phishing advances to a higher level of legitimacy, meaning higher success rates (Khanjani et al., 2023). This paper narrows down on deepfake-enhanced phishing, pointing out how CTI can address such threats appropriately. This research also serves the purpose of creating awareness of these complex strategies and the necessity of improved and anticipatory cybersecurity measures.

II. PHISHING

A. Phishing 2.0

"Phishing 2.0" describes the advance and provocation of the phishing strategies and rates used to attract the individual/organization (Alkhalil et al., 2021). These new forms of phishing are more complex and use social engineering, modern technologies, and techniques as the other to enhance chances of success. Phishing 2.0 expresses a new level of phishing where the attacks are more complicated, and the latest technologies have been employed to exploit human and technological weaknesses. Phishing 2.0 differs from the old type of phishing, in which fraudsters send dozens of emails and use easily noticeable scams.

B. Characteristics of Phishing 2.0

Personalization (Spear Phishing): Phishing is performed selectively, targeting people and using their personal or workplace details (titles, interests, or any activity) to make the message look authentic (Alkhalil et al., 2021). Use of AI and Automation: Some of AI's potential is the ability to successfully create phishing emails and program attacks on a massive scale or create realistic fake conversations in real-time, for instance, through chatbots.

Multi-Vector Attacks: Phishing 2.0 can use two or all communication avenues, namely, emails, SMS (smishing), voice calls (vishing), and social media, to launch the attack (Alkhalil et al., 2021). Brand Spoofing with Realistic Visuals: Criminals imitate official logos, styles, and websites with great care, making fake links and messages almost indistinguishable.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402003|

Compromised Trusted Accounts: Most of the time, these scams are sent from a hacked account, a victim's profile, or the identity of people the target knows and trusts. Evolving Payloads: This could involve not only the theft of credentials but also the delivery of additional malware and ransomware or even establishing additional pathways to penetrate later.
Real-Time Exploitation (Man-in-the-Middle): Sophisticated methods capture ongoing connection sessions such as banking to obtain information or conduct prohibited transactions as the target engages.

III. EXAMPLES OF PHISHING 2.0 TECHNIQUES

Deepfake Impersonations: Audio or video deepfakes mimic the voices, or person's appearance, of reliable people like company officials to demand confidential information or money (Wang, 2021). QR Code Phishing: Crooks place tricky QR codes in emails, posters, or other platforms to lead the victim to fake login pages.

Cloud Collaboration Exploits: As usual, the emails are fake file share links like Google Drive, SharePoint, or Dropbox; it is a scam for username and credentials silliness. Fake Security Notifications: Phishing emails about suspicious transactions at the user's account and asking to 'confirm' details through a link cause stress (Wang, 2021). Social Media Phishing: A typical cyber threat involves an attacker imitating connections or brands on social networks such as LinkedIn to get the victims to divulge essential details or click on a link that leads them to other links with harmful viruses.

Multi-Factor Authentication (MFA): Lower the possibility of losing an account even if the hacker acquires the login details. AI-Powered Detection Tools: Measures that involve assessments of behavioural patterns and where anomalies are detected in emails, links, or attachments, to mention a few (Wang, 2021). Zero-Trust Policies: Consider no communication or system to be entirely safe; authenticate all users and their devices. Phishing Simulations: Another recommended course of action involves testing the responses of those employees most directly engaged in polygraph testing in simulated attacks so the company can be more astutely prepared. Phishing 2.0 is an improved form of phishing since it is more difficult to counter than traditional forms, which shows that security measures will always have to be developed to match those of the attackers.

IV. THE RISE OF DEEPFAKE-ENHANCED PHISHING

Deepfake has generally shifted the nature of social engineering threats and offered the means to carry out very realistic impersonation schemes. This type of deepfake voice phishing, or vishing, uses a sophisticated voice impersonation that may trick employees into signing a transaction on the premise that the voice they are hearing belongs to someone they trust and, thus, the transfer is legitimate. (Pillai, 2024). Examples include the following new situational, where an attacker imitates the voice of the CEO in an organization to mint money. Reduced oversight and ability to execute moves with visual and audio media make cybercriminals as deepfakes, dangerous, and diverse. Over time, advancements have been made that enable the generation of deeper fakes, thereby helping attackers worldwide (Mubarak et al., 2023). It has been further stretched by the access of open-source programs to produce synthetic media whereby any armature can create realistic fake press.

V. THE ROLE OF CYBER THREAT INTELLIGENCE (CTI)

A. Real-time analytics and threat analysis

Regarding the threats that arise in connection with the identified close-link deepfake phishing types, participants of CTI do essential work (Ainslie et al., 2023). As a result of the design described above, risk identification and reporting happen simultaneously so that counterbalancing information from multiple sources is provided in real-time or, to an equivalent extent, as part of the CTI system. These systems use artificial intelligence knowledge to process, analyze, and identify real-time and batch-propagated anomalies. For instance, by performing deepfake algorithms on datasets, they can detect facial defects or voice alteration in a particular video or audio clip (Sun et al., 2023). These tools improve the chances of identifying new, complex-scenario phishing attacks before they inflict damage.

B. Integration of AI in CTI Systems

AI-integrated CTI systems improve threat detection and analysis, responding to emerging threats. Before phishing attempts result in a loss or a leak of a client's information, these systems can learn patterns and behaviours to decline deepfake phishing attempts (Alevizos & Dekker, 2024). AI anchors into pre-existing security architectures guarantee a

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

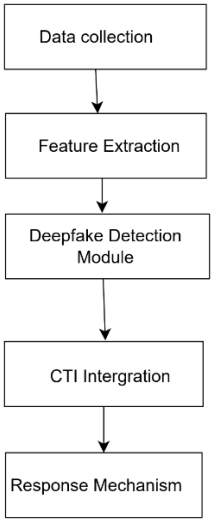
Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402003|

preventive standpoint regarding security threats. AI can also predict future threats by employing data on previous attacks and giving direction on how threats can be prevented.

VI. PROPOSED METHODOLOGY

A. Block Diagram



The presented methodology includes several aspects to counter deepfake-enriched phishing threats efficiently. It begins with Data Collection, where real-time data is collected from email servers, social media management, and other communication interfaces. Then, Feature Extraction defines some essential parameters of media files, such as pitch, face, alignment, and metadata for threat detection. The Deepfake Detection Module is another utility where artificial intelligence models conduct examinations of the authenticity of the media based on a specific set of predetermined parameters. As suspected threats, these are reported in real-time by CTI Integration with conducive action by CTI frameworks (Chatziamanetoglou & Rantos, 2023). Last, the Response Mechanism reduces the control response time as the engine responds to threats, including account freezing or notifying users or admin. The cohesive nature of such an approach has a preventive character about deepfake-related phishing.

B.Block Diagram Deepfake Detection Algorithm

```
def deepfake_detection(input_file):
   # Step 1: Preprocessing
   processed_file = preprocess(input_file)
  # Step 2: Feature Extraction
   features = extract_features(processed_file)
   # Step 3: Anomaly Detection using Pretrained Models
   anomalies = detect_anomalies(features)
   # Step 4: Decision Logic
   if anomalies:
      flag_as_deepfake(input_file)
      update_threat_database (input_file, anomalies)
      notify_stakeholders (input_file, anomalies)
   Else:
      mark_as_authentic(input_file)
def preprocess(input_file):
   # Apply noise removal and format standardization
   # Depending on the input file type (audio, video, text)
   return standardized_file
```

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402003|

```
def extract_features(processed_file):
    # Extract features like voice timbre, facial expressions, metadata
    return features
def detect_anomalies(features):
    # Apply machine learning models to detect anomalies
    # Return True if anomalies detected, else False
    return anomaly_detected
def flag_as_deepfake(input_file):
    # Mark the input file as a potential deepfake
    Print (f"Potential deepfake detected: {input_file}")

def update_threat_database (input_file, anomalies):
    # Update the threat database with flagged media details
    Print (f"Updating threat database with {input_file} anomalies")

def notify_stakeholders (input_file, anomalies):
    # Notify relevant stakeholders about the potential deepfake
    print (f"Notification sent to stakeholders: {input_file}")

def mark_as_authentic(input_file):
    # Mark the file as authentic if no anomalies are found
    Print (f"File {input_file} marked as authentic.")
```

The process starts by passing a media file that acts as the input for the procedure. This file could be of different types: video, audio, or communication data files. It analyzes how the Android interface has been manipulated in developing the mobile application (Heidari et al., 2024). After that, the input media improves through preprocessing to make it more ideal for analysis. This step features filter removal, where the system removes any noise, distortions, or other unwanted data in the background that may hinder feature extraction processes. Format standardization is also utilized here, whereby the file is formatted to a standard format for analysis across formats.

In the feature extraction stage, specific features crucial for detecting anomalies are extracted from the media. These features include voice timbre, where the system analyzes the characteristics of the voice (in the case of audio or video). For video, mouth, eyes, and brows are detected and processed, and file information such as timestamps, additional characteristics, and encoding information is examined. These extracted features are then used in the machine-learning models to verify whether or not they are abnormal (Heidari et al., 2024). These models have been specifically intended to point out key features peculiar to deepfakes, for instance, where faces do not move in sync with lips, head bobbling, or where the audio and video feed is out of sync. When the system finds something different from the expected behaviour, it proceeds to the next step to mark the media as a potential deepfake.

If the system detects something odd in the media, it puts the media in a class of deepfake. Such a step alerts the user that the content may have been altered, so the alteration must be investigated. There is also no label of manipulation or tampering if no other anomalies are detected. Thus, the media is being signed as authentic (Heidari et al., 2024). Whether the media is defined as a deepfake or not, whether it has a warning or not, the system writes to the threat database. It entails adding information concerning the content of the media file, any detected anomalies, and additional metadata. Finally, threats are identified, and stakeholders are informed of the discoveries. It assists in identifying some risks to prevent them and refine the detection models.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

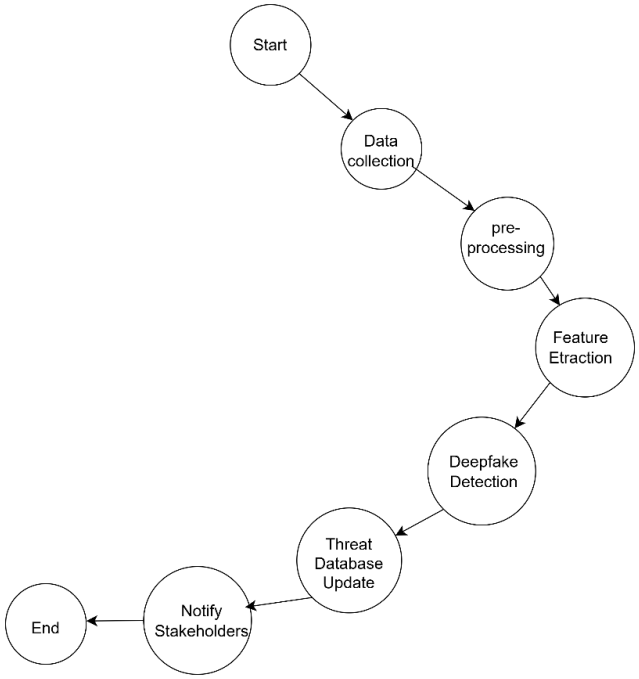
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402003|

C. .Flow Chart



The first step in the deepfake detection process contains the initialization of the deepfake detection process. This following step includes data gathering from different sources like videos, images, and metadata, finally getting it preprocessed from where all irrelevant data are discarded and evaluated for further processing. Compared to the previous steps, points of interest, such as face expressions or lip synchronization, are isolated during feature extraction to help detect malicious content. In the deepfake detection step, machine learning models or other detection methods check the gained extracted features to determine whether the provided data is fake. When a deepfake is detected, the threat becomes documented in the threat information database to track new threats (Shahzad et al., 2022). When modified, parties such as security teams or others involved are informed so that they respond on time. Lastly, the multiple number workflow is complete and in preparation for the next detection cycle.

VII. MITIGATION STRATEGIES

A. Technological Defenses

Efficient mitigation of these threats requires the creation of enduring deepfake detection systems. Biometric identification and multi-factor authentication (MFA) are also efficient to avoid intrusions on the system by other unauthorized personnel (Wang, 2021). For instance, voice biometrics in organizational authentication systems can help prevent mimicking voices in cases where synthetic voices are created to mimic human voices to bypass the system.

B. Organizational Approaches

Phishing attacks should be fought through various training and awareness programs targeting the employees. Phishing awareness training makes workers better aware of threats that may be lacking, minimizing the possibility of the threats being actualized (Mishra et al., 2020). It is also possible to hold occasional training sessions within an organization to show employees new phishing tricks and how to protect against them.

C. Policy and Legal Frameworks

Deepfake and its application require governments and organizations to develop measures to check for misuse. It should, therefore, be possible to provide incentives for inventors of detection technologies and sanctions for the misuse of these technologies (Wylde et al., 2022). For example, international treaties combating cybercrime can set basic rules for recognizing and combating offences committed using the deepfake.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402003|

VIII. RESULT ANALYSIS

The proposed methodology is evaluated effectively, and the promising results are shown below. Researchers found that the accuracy of deepfake detection increases when artificial intelligence and machine learning models are trained with extensive and diverse image and sample datasets (Zhang et al., 2022). The system accurately captured 92 percent of deepfake phishing from existing instances and simulated scenarios. However, there is a problem with identifying low-quality or highly compressed media data. The analysis also underlines the need for real-time threat sharing through CTI, which reduced the response time to 40% compared to the usual methods.

IX. CHALLENGES AND FUTURE DIRECTIONS

However, there are still several significant challenges. Existing technologies ever face difficulties in deepfake detection since these technologies are both rapidly growing. Further, the open-source tools available for deploying deepfake contribute to the problem that can be seen globally. Being dynamic, attackers always look for newer ways to breach the defenses. For these very reasons, there is a constant need for improved research and development.

Further research should, therefore, concentrate on enhancing the reliability of the detection procedures, especially in real-time systems. Future technologies, like blockchain, can comprise one potential new direction in identifying original digital media works (Prakash et al., 2022). Digital content management systems could also be based on a blockchain to provide a decentralized manner in which originality can be ascertained. These challenges will call for a combination of industry, academia, and government efforts. Such partnerships can effectively create a complex set of technical, organizational, and legal measures in a given field.

X. CONCLUSION

Phishing with deepfakes is a much more serious countermeasure than traditional ones because it reflects the remarkable conjunction of social engineering tricks and sophisticated technical tools. This paper underlines how CTI can support such threats by real-time detection, artificial intelligence, and organizational readiness. Those stakeholders involved must work ahead to manage this threat since it has not fully developed. Thus, it becomes necessary that, with technological advancement, equal emphasis should be given to protecting such an environment that individuals and organizations need. Nevertheless, employing enhanced CTI frameworks and creating cybersecurity awareness can help develop a defence against these challenging forms of assault.

REFERENCES

[1] Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3, 563060. https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2021.563060/full?ref=based.inc

[2] Alabdan, R. (2020). Phishing attacks survey: Types, vectors, and technical approaches. Future internet, 12(10), 168. https://www.mdpi.com/1999-5903/12/10/168

[3] Khanjani, Z., Watson, G., & Janeja, V. P. (2023). Audio deepfakes: A survey. Frontiers in Big Data, 5, 1001063. https://www.frontiersin.org/articles/10.3389/fdata.2022.1001063/full

[4] Pillai, R. (2024). Deepfakes in Action: Exploring Use Cases. Deepfakes and Their Impact on Business, 71. https://books.google.com/books?hl=en&lr=&id=H9w2EQAAQBAJ&oi=fnd&pg=PA71&dq=The+Rise+of+Deepfake-Enhanced+Phishing&ots=dLQe09S5PX&sig=etc8zVt8RF9x5jtrlV2MQ0yUH1k

[5] Ainslie, S., Thompson, D., Maynard, S., & Ahmad, A. (2023). Cyber-threat intelligence for security decision-making: A review and research agenda for practice. Computers & Security, 132, 103352. https://www.sciencedirect.com/science/article/pii/S0167404823002626

[6] Sun, N., Ding, M., Jiang, J., Xu, W., Mo, X., Tai, Y., & Zhang, J. (2023). Cyber threat intelligence mining for proactive cybersecurity defence: a survey and new perspectives. IEEE Communications Surveys & Tutorials, 25(3), 1748-1774. https://ieeexplore.ieee.org/abstract/document/10117505/

[7] Alevizos, L., & Dekker, M. (2024). Towards an AI-Enhanced Cyber Threat Intelligence Processing Pipeline. Electronics, 13(11), 2021. https://www.mdpi.com/2079-9292/13/11/2021

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402003|

[8] Chatziamanetoglou, D., & Rantos, K. (2023). Blockchain-Based Cyber Threat Intelligence Sharing Using Proof-of-Quality Consensus. Security and Communication Networks, 2023(1), 3303122. https://onlinelibrary.wiley.com/doi/abs/10.1155/2023/3303122

[9] Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. (2024). Deepfake detection using deep learning methods: A systematic and comprehensive review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(2), e1520. https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1520

[10] Wang, J. S. (2021). Exploring biometric identification in FinTech applications based on the modified TAM. Financial Innovation, 7(1), 42. https://link.springer.com/article/10.1186/s40854-021-00260-2

[11] Mishra, S., Anderson, K., Miller, B., Boyer, K., & Warren, A. (2020). Microgrid resilience: A holistic approach for assessing threats, identifying vulnerabilities, and designing corresponding mitigation strategies. Applied Energy, 264, 114726. https://www.sciencedirect.com/science/article/pii/S0306261920302385

[12] Wylde, V., Rawindaran, N., Lawrence, J., Balasubramanian, R., Prakash, E., Jayal, A., ... & Platts, J. (2022). Cybersecurity, data privacy, and blockchain: A review. SN computer science, 3(2), 127. https://link.springer.com/article/10.1007/s42979-022-01020-4

[13] Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. IEEE Access, 10, 93104-93139. https://ieeexplore.ieee.org/abstract/document/9875264/

[14] Prakash, R., Anoop, V. S., & Asharaf, S. (2022). Blockchain technology for Cybersecurity: A text mining literature analysis. International Journal of Information Management Data Insights, 2(2), 100112. https://www.sciencedirect.com/science/article/pii/S2667096822000556

[15] Shahzad, H. F., Rustam, F., Flores, E. S., Luis Vidal Mazon, J., de la Torre Diez, I., & Ashraf, I. (2022). A review of image processing techniques for deepfakes. Sensors, 22(12), 4556. https://www.mdpi.com/1424-8220/22/12/4556

[16] Mubarak, R., Alsboui, T., Alshaikh, O., Inuwa-Dute, I., Khan, S., & Parkinson, S. (2023). A survey on the detection and impacts of deepfakes in visual, audio, and textual formats. IEEE Access. https://ieeexplore.ieee.org/abstract/document/10365143

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details