A Comparative Study of Deepfake Threat Landscape in Surface Web vs. Dark Web

Shaikh Junaid Ahmad[1], Dr. P.A. Kadam[2]
Research Scholar, School of Computational Sciences, SRTMU Nanded[1]
Assistant Professor, Institute of Technology & Management (SSBES' ITM), Nanded[2]
shaikh.junaid24@gmail.com[1], puru.kadam@gmail.com[2]

Abstract

Deepfake technology, initially created for entertainment and experimental AI applications, has rapidly become a significant concern in cybersecurity and digital trust. This paper compares the evolving threat landscape of deepfakes across two layers of the internet: the surface web and the dark web. On the surface web, deepfakes are commonly used in political misinformation, viral hoaxes, and non-consensual pornography, often spread via social media and public forums, where some moderation is in place (Whittaker et al., 2023). In contrast, the dark web facilitates a more hidden, commercialized use of deepfakes for criminal purposes, such as extortion, identity fraud, and the sale of synthetic media tools (Abbas et al., 2025). These activities thrive due to the high degree of anonymity and lack of regulation (Dami, 2022). By analysing user intent, accessibility, distribution, and legal oversight, this study highlights the stark contrast between surface-level and deep-layer exploitation of synthetic media. The findings suggest that while surface web threats are more visible and subject to moderation, deepfakes on the dark web represent a more severe, organized risk that is harder to trace or mitigate (Sandoval et al., 2024). This calls for urgent interdisciplinary solutions that combine technological innovation with policy reform and international cooperation.

Keywords: Deepfake cybersecurity, Surface web vs dark web, Synthetic media threats, Digital trust andmisinformation, AI-enabled cybercrime.

1. INTRODUCTION

1.1 Background

Deepfakes are fake synthetic-form media content in videos and images that are created or altered by artificial means through deep-learning methods in particular, based on Generative Adversarial Networks (GANs). These programs are able to produce highly realistic yet completely manufactured visual and audio content that is difficult for a lay viewer to determine as real or fake. Initially unveiled as an item of technological interest in entertainment and visual effects productions, deepfakes have been increasingly becoming serious tools of deception (Dami, 2022). Ranging from impersonating famous personalities to producing non-consensual pornography, deepfakes have shown that they have the capability of twisting reality in false directions, manipulating perception, and compromising integrity in digital-media ecosystems. As these synthetic technologies advance, their use has split into different levels of the internet. The publicly accessible part of the internet, or the surface web, has seen extensive use of deepfakes on social platforms like Twitter, YouTube, and TikTok. These have predominantly been about misinformation campaigns, parody videos, and viral hoaxes eliciting mass public interest (Whittaker et al., 2023). Deepfakes in the dark part of the internet available only through encrypted networks such as Tor have become commodities for sale as part of illicit offerings ranging from identity impersonation to digital extortion kits (Abbas et al., 2025). The extreme difference in visibility, intent, and regulation in these two domains is an urgent area of scholarly and policy interest.

1.2 Objectives and Goals

This research seeks to methodically examine, compare, and dissect the environment of deepfake threats on the surface web and dark web. The principal objectives are

- Determination of major types and distribution channels of deepfakes in both fields.
- Examining user motivations that include fame, influence, or criminal motives for producing and disseminating deepfakes.
- Reviewing the extent of regulation and detection tools employed in order to counter deepfake attacks.
- Emphasizing societal and cybersecurity implications of these threats in open and hidden online environments.
- To examine financial losses through deepfake in different domain. Through an analysis of these elements, the research aims to enlighten interdisciplinary approaches bridging technological, legal, and policy solutions for countering deepfake abuse.

1.3 Literature Review: Scholarly work on deepfakes has increasingly increased over the last five years in proportion to the growing breadth and influence of the technology. The majority of research on the surface web deals with the public release of deepfakes on social media websites, video platforms, and online forums. Deepfakes are usually political in content and are employed to disseminate false information, influence elections, or defame people. According to Whittaker et al. (2023), deepfakes have permeated popular culture and normalized synthetic media in a manner that has not been accompanied by ample discussion of ethical risks.

Conversely, the dark web has garnered lesser scholarly interest but is increasingly of interest to security experts. Abbas et al. (2025) and Sandoval et al. (2024) report on the commodification of deepfakes in encrypted websites for sale for use in scams and frauds and even orchestrated cyber-attacks. These platforms ensure high degrees of anonymity and encryption for those who use them, making detection and law enforcement particularly challenging. According to Dami (2022), in contrast to the surface web that operates under platform regulation and relatively moderate amounts of legal oversight, the dark web exists as a largely unregulated environment in which malicious actors have a field day.

All prior research underscore a shared thread in that whereas surface-web deepfakes are more overt and socially destabilizing, dark-web deepfakes pose a deeper, systemic risk in their organized and commodified form. This dualism necessitates the use of a comparative method in order to address the multilateral threat which synthetic media has come to pose throughout the internet.

2. Methodology

This research employs a qualitative comparative method for analysing use of deepfakes on the dark and surface web. The research is based on peer-reviewed literature, white papers, and cybersecurity reports, and its main sources include Abbas et al. (2025), Dami (2022), Sandoval et al. (2024), and Whittaker et al. (2023). Analysis is organized in five major dimensions:

1. **Types of Deepfake Content**: Deepfakes can be categorized based on form and use (e.g., political disinformation, identity impersonation).

2. **Distribution Methods**: Examining how deepfakes spread across platforms like YouTube, TikTok, and Tor marketplaces.

3. **User Motivatedness and Anonymity**: Examining user motivations for creating deepfakes and anonymity in the dark web.

4. **Detection and Regulation**: Detection technologies and regulation models in both the surface and dark web.

5. **Societal Risks**: Examining wider implications of cybersecurity, privacy, and digital trust from deepfakes.

The study is based on publicly available data collected from online threat intelligence sources and scholarly research, but not through direct access to illegal darknet channels.

3. Results & Discussion

An analysis of contemporary literature and cybersecurity reports reveals a significant escalation in the prevalence, impact, and societal consequences of deepfake technologies across the surface web and the dark web.

3.1 Deepfake Types and Distribution Channels

The volume and type of deepfakes have increased dramatically. As of 2024, more than 95,820 deepfakes have been discovered on major platforms, an increase of 550% since 2019 [18]. These videos are shared through both mainstream platforms (e.g., social media platforms, content sharing sites) and encrypted forums on the dark web. Surface web platforms are under user reporting and takedown mechanisms, while dark web enables unregulated sharing through forums, marketplaces, and P2P networks, due to which the persistence and virality of adverse content are enhanced.

3.2 Motivations behind Deepfake Creation

Deepfake creators often operate under diverse motivations. On the surface web, motives include **fame, satire, or social influence**. In contrast, dark web usage leans toward **malicious intent**, including extortion, political manipulation, and synthetic identity fraud. The **financial impact on victims** supports this distinction—**77% of affected individuals** report financial loss, with **7% losing between $10,000 and $15,000** (McAfee, 2023), indicating criminal exploitation as a dominant dark web driver.

3.3 Detection and Regulation

Detection efficacy varies significantly between humans and machines. Human detection remains unreliable, with an average **accuracy of only 57%** (Pew Research, 2023). In contrast, **AI-based systems** have shown **up to 84% accuracy** in spotting manipulated videos (Security Hero Report, 2023). Despite these advances, surface web platforms still face challenges due to the volume and speed of content sharing. On the dark web, the **absence of platform moderation and regulatory oversight** renders most detection tools ineffective or inapplicable.

3.4 Societal and Cybersecurity Implications

Proliferation of deepfakes has serious implications. 96% of deepfakes are pornography, and 98% of those include women as targets of attack, namely celebrities (Sensity, 2023). Apart from escalating gender-based online bullying, it also leads to psychological, reputational, and monetary harm. On the dark net, these risks are compounded by anonymity and longevity of exchanged content. The social harm extends from invasion of privacy, disinformation campaigns, and erosion of confidence in visual content.

3.5 Financial losses by deepfake

Deepfake incidents manifest differently across the surface web and dark web, with distinct patterns in reporting and use cases. Below is a breakdown based on available case studies and reports:

Surface Web Incidents

1. **Public Scams and Fraud**

- The Kerala deepfake fraud case (2022) involved a scammer impersonating a victim’s former colleague via WhatsApp, resulting in a ₹40,000 loss. This incident was publicly reported and investigated by local authorities [19].

- Palo Alto Networks identified **hundreds of domains** hosting deepfake scams in 2024, including fake investment schemes (e.g., "Quantum AI") and government giveaways. These campaigns targeted users in Canada, Mexico, France, and others via surface web platforms[20].

2. **Political Disinformation**

- A 2023 deepfake video falsely showed Singapore’s Prime Minister endorsing a cryptocurrency platform, causing public confusion[20].

- In 2024, synthetic media of political figures like Joe Biden and fabricated images of explosions at the Pentagon impacted elections and financial markets[20].

3. **Business and Hiring Attacks**

- The 2024 KnowBe4 incident involved a deepfake candidate bypassing remote hiring checks, highlighting vulnerabilities in corporate verification processes[22].

- 49% of companies reported encountering audio/video deepfakes in 2024, up from 29% in 2022[21].

4. **Quantitative Trends**

- Deepfake fraud attempts surged by **2,137%** between 2020 and 2023 on the surface web2[21]

- Palo Alto Networks observed scam domains averaging **114,000 visits each** before takedown.

Dark Web Activity

1. **Tool Distribution and Services**

- Kaspersky's 2024 analysis revealed a thriving marketplace for deepfake creation tools on darknet forums. Services include:

- Face-swapping software (e.g., Swapface) priced at **$0–$249/month**[21].

- Custom deepfake videos for identity fraud, bank scams, and disinformation campaigns.

2. **Data Trading**

- Cybercriminals trade stolen personal data (e.g., social security numbers, bank details) to train deepfake models. For example, the Homeland Security report describes attackers harvesting social media content to create voice clones for financial fraud [21].

3. **Monetization of Attacks**

- Dark web forums host tutorials for deploying deepfakes in ransomware, crypto currency scams, and credential theft. The **$25 million Arup deepfake heist** (enabled by dark web-sourced tools) exemplifies this trend.

3.6 Comparison with Prior Work

The findings align with studies by **Dawson (2021)**, **Abbas et al. (2025)**, and others, which document a converging concern over the accelerated deployment of deepfakes. This study contributes by drawing a **comparative perspective**, emphasizing that while both surface and dark web environments face deepfake threats, the **nature, intent, and impact differ substantially** between them. It reinforces the need for **tailored detection mechanisms, legal reform, and awareness initiatives** based on platform type.

3.7 Interpretation

- Deepfakes on the **surface web** are more visible but are regulated by platform policies and legal oversight.

- The **dark web** presents more profound, long-term challenges due to its lack of regulation and the difficulty in detecting or mitigating malicious use.

3.8 Comparison with Prior Work

- The findings of this study align with previous research, such as that of **Dawson (2021)**, **Abbas et al. (2025)**, and others, reinforcing the thematic convergence around the growing risk posed by deepfakes across both the surface and dark web.

4. Comparative analysis

Aspect	**Surface Web**	**Dark Web**
Visibility	High; public domain with partial moderation	Low; accessible only through encrypted browsers (e.g., Tor)
Motivation	Political influence, social clout, scams	Financial extortion, organized crime, fraud kits
Regulation	Subject to platform policies and national laws	Largely unregulated; jurisdictional challenges
Distribution	Social media, forums, streaming platforms	Darknet forums, encrypted chats, marketplaces
Detection	Growing use of AI-based tools;	Extremely difficult; anonymity and

	some success in takedown	decentralization hinder detection
Impact	Social disruption, misinformation, individual financial loss	High-value fraud, geopolitical manipulation, identity exploitation

Table 1: Comparative analysis of surface web and darkweb

5. Conclusion

5 .1 Summary

The deepfake threat is multi-layered: the surface web highlights widespread misuse in public domains, while the dark web reveals deeper, organized, and criminal exploitation.

5.2 Implications

- There is a pressing need for more robust, cross-platform detection systems.
- International policy collaboration is essential to address anonymity-driven misuse, particularly on the dark web.

6. References

[1] L. Dami, “Analysis and conceptualization of deepfake technology as cyber threat,” *ResearchGate*, 2022.

[2] R. Abbas *et al*., “Harnessing Big Data Analytics for Advanced Detection of Deepfakes and Cybersecurity Threats Across Industries,” 2025.

[3] M. E. Dawson, “Cyber warfare: threats and opportunities,” 2021.

[4] J. Whitman *et al*., “Adversarial Machine Learning in Dark Web Threat Detection,” 2024.

[5] M. Anisetti *et al*., “Security threat landscape,” *CONCORDIA*, 2020.

[6] J. Zhang and D. Tenney, “The Evolution of Integrated APTs and Deepfake Threats,” 2023. [Online]. Available: HTML Paper.

[7] A. K. Tyagi *et al*., “Security and Possible Threats in Today’s Online Social Networking Platforms,” *Wiley*, 2024.

[8] R. P. Shukla *et al*., “Combatting Deepfake Threats in India: A Data-Driven Approach,” *IGI Global*, 2024.

[9] L. Whittaker *et al*., “Mapping the Deepfake Landscape for Innovation,” *Technovation*, 2023.

[10] M. P. Sandoval *et al*., “Threat of Deepfakes to the Criminal Justice System: A Systematic Review,” *Springer*, 2024.

[11] C. Chipeta, “Deepfake Statistics (2024): 25 New Facts for CFOs - Eftsure,” *Eftsure*, Jul. 12, 2024. [Online]. Available: https://eftsure.com/statistics/deepfake-statistics.

[12] A. Lewis *et al*., “Deepfake Detection with and without Content Warnings,” *Royal Society Open Science*, vol. 10, no. 11, Nov. 1, 2023. [Online]. Available: https://doi.org/10.1098/rsos.231214.

[13] Sensity AI, “Deepfake Report: Threat Landscape and Trends,” 2023. [Online]. Available: https://sensity.ai

[14] Sensity.ai, “Top Deepfake Detection Solution | New AI Image Detection,” Jan. 29, 2023. [Online]. Available: https://sensity.ai/deepfake-detection/

[15] Pew Research Center, “The Rise of Deepfakes and the Public’s Struggle to Detect Them,” 2023. [Online]. Available: https://www.pewresearch.org

[16] McAfee Corp., “The Truth About Deepfakes: Consumer Impacts Survey,” 2023. [Online]. Available: https://www.mcafee.com

[17] Security Hero, “AI in Cybersecurity 2023 Report: Deepfake Detection Benchmarks,” 2023. [Online]. Available: https://www.securityhero.com

[18] Shalwa, “AI-Generated Deepfakes: Statistics, Market Growth, Trends,” *Artsmart.ai*, Jan. 27, 2025. [Online]. Available: https://artsmart.ai/blog/ai-generated-deepfakes-statistics/

[19] Indian Cyber Squad, “Case Study: Kerala’s First Deepfake Fraud,” Nov. 27, 2023. [Online]. Available: https://www.indiancybersquad.org/post/case-study-kerala-s-first-deepfake-fraud

[20] Keepnet Labs, “Deepfake Statistics about Cyber Threats and Trends 2024 - Keepnet,” Apr. 16, 2024. [Online]. Available: https://keepnetlabs.com/blog/deepfake-statistics-and-trends-about-cyber-threats-2024

[21] IProov, “The KnowBe4 Deepfake Incident - a Wake-up Call for Remote Hiring Security,” Aug. 20, 2024. [Online]. Available: . https://www.iproov.com/blog/knowbe4-deepfake-wake-up-call-remote-hiring-security

[22] Kaspersky Team, “How Real Is Deepfake Threat?” *Kaspersky.co.in*, May 11, 2023. [Online]. Available: https://www.kaspersky.co.in/blog/deepfake-darknet-market/25663/. Accessed: May 10, 2025