Shufti

REPORT

Outsmarting the Deepfake Threat to Identity Trust

Outsmarting the Deepfake Threat to Identity Trust

Digital trust is at a critical inflection point. What once seemed like a fringe novelty, deepfake technology has advanced rapidly from its early use in academic research and filmmaking, becoming a powerful tool for deception. With the rise of generative adversarial networks (GANs) and accessible deepfake creation tools, fraudsters can replicate faces, voices, and behaviors with stunning accuracy.

The result: an identity fraud landscape that's no longer constrained by geography, technical skill, or time zones.

Deepfakes are already reshaping fraud across sectors. In financial services, gaming, and insurance, synthetic identities and AI-generated media are being used to spoof KYC interviews, hijack accounts, and impersonate executives.

Biometric deepfakes now account for up to **40% of identity attacks**, with new attempts occurring every five minutes.[1] More so, **account takeover (ATO) and identity fraud incidents** due to generative AI and deepfakes **increased by 244%.**[2]

Legacy verification methods, including manual ID checks, are increasingly vulnerable in this new threat environment. The financial impact is mounting, with deepfake-enabled fraud driving significant losses from identity theft, account takeover, impersonation and business email compromise. As the line between real and artificial becomes increasingly blurred, organizations need to reassess and modernize their identity verification frameworks to maintain trust and security.

Shufti is on the frontlines of this battle, deploying a robust, multi-layered defensive scheme against deepfakes. By utilizing cutting edge AI techniques, behavioral biometrics, and continuous model training, Shufti provides real-time, scalable protection against the ever growing onslaught of attacks. Plus, full ownership and control of our own tech stack enables us to release as many as ten updated detection layers a month, far more than providers that are dependent on third-party infrastructure can in the same amount of time.

This report explores the rise and impact of deepfake fraud, how it's being deployed, and what businesses must do to stay ahead. It also outlines how Shufti's proprietary systems combine liveness detection, multimodal validation, and adaptive risk scoring to combat synthetic media at scale. Our recommendations call for proactive adoption of proven detection tools, ongoing model improvement, and industry-wide collaboration to counter this rapidly evolving threat.

1 | Vakulov, Alex. "Deepfake Scams Are Stealing Millions-How to Spot One." Forbes, Forbes Magazine, 9 Mar. 2025.
2 | Shufti. "Deepfakes: AI Fraud Attacks Require Even Smarter AI Countermeasures. Now." Accessed 14 05 2025

The Evolution of Deepfake Fraud

1

From Novelty to Threat

Some of the earliest attempts at creating deepfakes can be traced back to the 1990s, when CGI artists and researchers at academic institutions attempted to render realistic human images.[3] However, the technology at the time could not match their ambitions, and the visuals they were able to create lacked consistency and realism. The real turning point came in 2014, when Ian Goodfellow, a doctoral student at the Université de Montréal, introduced the first GAN, a model able to generate highly realistic images through adversarial neural networks.[4]

Prior to GANs, face manipulation was primarily driven by autoencoders, which compressed and restructured image data. While these were a step forward, the images often appeared smooth and unnatural. The introduction of GANs changed everything. In 2017, the term "deepfake" was coined and entered into the mainstream through Reddit forums, serving as a catalyst to the creation of tools that could convincingly mimic human appearances

.

Early applications of deepfakes in the world of fraud were crude and relied on single-modal deception, usually face-swapping or voice synthesis, and often fell short when it came to believability. Fraudsters tended to target static assets like passwords and government IDs, relying heavily on human error rather than machine precision.

Today's deepfakes, however, are dynamic, coordinated, and capable of true deception.

They target remote verification systems and exploit biometric vulnerabilities, making them exceedingly hard to detect and far more dangerous.

3 | "A Brief History of Deepfakes." Reality Defender - Enterprise-Grade Deepfake Detection, 1 June 2024.

4 | Goodfellow, Ian J., et al. "Generative Adversarial Networks." arXiv.Org, 10 June 2014.

The Latest Generation of Deepfakes

Modern synthetic media has transformed in recent years from single-modal manipulation to highly advanced, multimodal simulations. These deepfakes incorporate facial expressions, voice patterns, and behavioral cues into one lifelike persona capable of passing through traditional verification systems. They feature highly accurate lip-syncing and eye movement rendering, using diffusion models and GANs to generate videos and digital masks that are capable of mimicking genuine human emotion and cognition.

One of the most dangerous breakthroughs in the world of deepfakes is the evolution of real-time generation. These systems can adapt and respond dynamically during remote KYC interviews or video calls, making them especially effective in impersonating executives or fooling onboarding systems. Unlike older versions that struggled with blinking rates, both too much and not enough, or expression coherence, modern models are able to achieve naturalistic blinking patterns and nuanced facial dynamics, further blurring the line between synthetic and authentic.

The proliferation of open-source deepfake tools in recent years have significantly democratized their creation and evolution. While many are used for entertainment, their accessibility allows entry-level fraudsters to conduct low-risk attacks for little to no money. In contrast, proprietary black-hat models, built specifically for fraud, exhibit dangerous levels of multimodal accuracy. These models are often used by criminal organizations to conduct large-scale synthetic fraud, producing hundreds of synthetic identities in a short period of time that are all capable of bypassing verification systems.

Actors Behind the Technology

The rise of fraud-as-a-service (FaaS) has lowered the barrier to entry for these high-level fraud schemes even further. Operating through marketplaces on platforms like Discord and Telegram, places accessible to just about anyone, these vendors sell specialized deepfake kits or offer their services for under $500, a mere fraction of the potential payouts from a successful scheme. These kits can include everything from voice-cloning software to pre-packaged synthetic identity templates.

Increasingly, deepfake tools have also been weaponized by state-sponsored actors and organized crime groups.

Entities like North Korea's Lazarus Group have begun to employ synthetic media to extract intelligence, manipulate global communications, and destabilize digital infrastructure.[5] This meeting of technological advancement and hostile intent has created a global threat landscape that shows no signs of slowing down any time soon.

5 | Abyazov, Emir. "Lazarus Group Targets Crypto Leaders with Deepfake Zoom Attacks." Coinpaper, Coinpaper, 21 Apr. 2025.

Attack Vectors and Deployment Channels

2

Primary Use Cases in Fraud

Synthetic identities have found fertile ground in digital onboarding environments, in no small part due to the COVID-19 pandemic and the hasty shift to completely remote verification for many industries. In the remote KYC/AML onboarding process, deepfakes can spoof video identities, clone voices, and present synthetic documents, all designed to pass visual verification inspection. Fraudsters exploit these vulnerabilities to access everything from bank accounts to crypto wallets.

The job market has also become a significant target for synthetic identity fraud. AI-generated personas are now used in recruitment scams where fraudsters pose as legitimate job seekers to secure employment in order to access internal systems, siphon payroll funds, or set up to support future attacks. These identities often come with generated faces, stolen credentials, and manipulated job searching profiles.

Another key threat vector now involves impersonation attacks in business email compromise (BEC) schemes. In these cases, deepfakes are used to impersonate executive voices or video appearances in order to convince employees to authorize high-value transfers. Unlike phishing, these scams are highly targeted, difficult to detect in real time, and rely more on the realism of the deepfakes than on human error.

Deepfake-enabled social engineering is also infiltrating fintech platforms like payment platforms, crypto exchanges, and neobanks. Fraudsters now use voice spoofing to impersonate representatives, tricking users into transferring funds to fraudulent accounts or revealing account credentials. As digital finance interfaces grow more reliant on online verification, fraudsters' ability to deceive expands significantly.

Deepfake-as-a-Service (DaaS) Ecosystems

Communities on platforms like Discord and Telegram have capitalized on the rising demand of deepfakes and commercialized synthetic fraud. DaaS providers offer "custom deepfakes in minutes," allowing users to create synthetic personas or voice clones with minimal technical experience. These virtual storefronts function like other e-commerce sites, complete with pricing tiers and customer service, often utilizing cryptocurrency as payment to enhance anonymity.

Even more dangerously, API-based services now let users upload images and text to generate realistic video personas. Voice cloning services can take as little as 30 seconds of audio and produce astonishingly realistic results. These APIs can automate deepfake generation for KYC evasion, making fraud scalable and accessible to just about anyone with an internet connection.

Real-World Case Studies

One of the first widely reported cases of deepfake fraud was in 2019, just five years after the development of GAMs. It involved the theft of €220,000 from a UK based energy company after fraudsters called the CEO and impersonated the firm's parent company's chief executive.[6] The funds were transferred to an account in Hungary, then forwarded to an account in Mexico, and continued as such, making the recovery efforts incredibly difficult.

In another instance, a Hong Kong-based employee at a multinational firm was tricked into transferring HK$200 million after receiving a credible email and joining a video call where scammers used deepfakes to impersonate the company's UK-based CFO and several colleagues.[7] Believing they had spoken to the company's CFO, the employee executed multiple transfers across five bank accounts, and the scam was only uncovered after the real CFO raised concerns days later. This incident is generally known as the first instance of multi-person deepfake video fraud.

A third case involves a highly sophisticated cybercrime run operation where young men would apply for jobs under the guise of being American while actually working for the North Korean government.[8] These applicants were able to simulate real documents during job interviews by using green screens and projected IDs while utilizing digital deepfake projected masks and voice synthesis to hide their true identities. These applicants were hired, infiltrated company systems, and attempted to reroute funds to accounts based in North Korea. While several individuals have been arrested and sentenced for their involvement in this scheme, it is estimated that there are still thousands of operatives infiltrating companies around the world.

6 | Damiani, Jesse. "A Voice Deepfake Was Used to Scam a CEO out of $243,000." Forbes, Forbes Magazine, 3 Sept. 2019.
7 | Burt, Andy. "Finance Employee Defrauded for $25M by Deepfake CFO." CFO.Com, 5 Feb. 2024.
8 | Johnson, Bobbie. "Your Favorite New Coworker Is an AI-Enhanced Operative from North Korea." Wired, Conde Nast, 1 May 2025.

Measuring the Scope of Damage

3

Quantifying the Impact

The scale of financial loss tied to synthetic identity fraud around the world is staggering and virtually impossible to fully calculate. According to the U.S.'s FBI Internet Crime Complaint Center (IC3), there were over $2.7 billion in reported BEC-related losses in 2022 alone.[9] **Synthetic identities now account for an estimated 10-20% of credit-related fraud[10] and cybercrime is projected to cost $10.5 trillion globally in 2025.[11]**

Deepfakes enable fraud at a speed and scale that would have been unimaginable just ten years ago. Automation allows for hundreds of onboarding attempts per day, with some studies reporting that a deepfake attack occurs approximately once every five minutes.[12] With minimal inputs and downtime, fraudsters can create fully formed digital identities and pass them through traditional verification systems with shockingly high success rates.

Beyond the immediate impact of financial losses, the reputational damage that comes from deepfake attacks can cripple companies. Trust erosion deters new users and can lead to regulatory scrutiny, especially when institutions are seen as soft targets. Cleanup and recovery operations also drain resources, diverting attention from growth-oriented initiatives.

9 | "2022 Internet Crime Report." Internet Crime Complaint Center, FBI. Accessed 7 May 2025.

10 | Richardson, Bryan, and Derek Waldron. "Fighting Back against Synthetic Identity Fraud." McKinsey & Company, McKinsey & Company, 2 Jan. 2019.

11 | Morgan, Steve. "Cybercrime to Cost the World $10.5 Trillion Annually by 2025." Cybercrime Magazine, 18 Nov. 2024.

12 | Opiah, Abigail. "Deepfake Attacks Now Occur Every Five Minutes, Entrust Report Warns: Biometric Update." Biometric Update | Biometrics News, Companies and Explainers, BiometricUpdate.com, 22 Nov. 2024.

Industries Most Affected

Naturally, financial services, fintechs, and neobanks are all prime targets for synthetic identity attacks. The high volume of virtual onboarding, automated transaction flows, and regulatory burdens make them ideal marks for fraudsters.

Gaming platforms, especially Web3-integrated environments, also face increasing threats as they rely heavily on blockchain-integrated platforms that offer a great deal of anonymity. Players will often verify their identity once to access crypto rewards, opening the door for fraudsters to clone accounts and continuously siphon rewards.

Insurance firms have also become a growing target in recent years, particularly when it comes to life insurance policies. Fraudsters will create fictitious identities by combining real personal information with fake details and then purchase a life insurance policy in order to collect on it later. Deepfakes are used to pass video interviews and project fake documents, making detection extremely difficult without advanced verification systems.

Detection Gap and Regulatory Urgency

Despite the rising threat, it is plain to see that regulation around deepfakes lags behind. While bodies like the FATF have issued high level guidance, there are currently few sufficient mandates that directly address the threat of real-time deepfakes. Biometric systems built before the deepfake boom are not properly equipped to identify synthetic anomalies, and the pace of legislation, like the 2024 AI Act in the EU and the DEEPFAKES Accountability Act in the U.S., is slow.

The detection gap is widening and the threat is only growing. Without multi-modal verification and behavioral modeling, traditional systems are blind to many of today's highly sophisticated threats. This lack of adaptation exposes companies to compounding risks that require immediate technological intervention.

Shufti's Multi-Layered Deepfake Defense Strategy

4

Foundations of Shufti's Detection Architecture

Shufti's defense architecture is built on the core understanding that detecting deepfakes in today's threat landscape requires more than just facial recognition or static liveness checks. It requires a layered, adaptive system that combines visual, behavioral, and contextual analysis to catch what generative AI is designed to mimic. Because we own and operate our entire tech stack, we're able to push detection updates in real-time, adding new layers as frequently as ten times a month, while other industry leaders are at the mercy of their third-party vendors.

Liveness detection serves as a critical first line of defense against impersonation and Shufti supports both active and passive modes depending on a client's risk environment. Active liveness detection prompts users to position their face within a designated frame, giving our systems a controlled environment to capture their image for verification. In passive mode, the system runs silently in the background, assessing light reflection patterns, skin texture, and depth to confirm genuine presence without requiring any user interaction.

To augment these checks, behavioral biometrics are used to track micro-movements such as blink frequency, pupil dilation, and lip movement. These signals are difficult to forge even with high-end deepfakes, providing an extra layer of confidence. For instance, unnatural stillness or erratic eye movements are common tells in AI-generated content and thus trigger further scrutiny by the system.

In addition, Shufti runs advanced forensic analysis on submitted images, using metadata and compression artifacts to verify that images or videos were captured directly by a physical camera, and have not passed through an interim editing step. By comparing the sensor noise pattern against a database of over 100,000 known camera signatures, we are able to confirm if the media was captured natively or has been post-processed, re-encoded, or otherwise manipulated.

Deepfake-Specific Enhancements

Deepfake attacks exploit the realism of synthetic media in order to bypass standard verification protocols. To address this, Shufti applies specialized detection models trained to identify patterns that are unique to AI-generated content.

Our deepfake classifiers analyze skin tone gradients, texture consistencies, and shadow continuity using convolutional neural networks trained on more than 50,000 synthetic documents.

These help detect so-called “uncanny valley” anomalies in portraits like asymmetrical facial features and unnatural skin tones.

What distinguishes our approach is the use of targeted detection layers built specifically for GAN-based and diffusion-based threats. Instead of relying solely on general biometric markers, our models are trained with synthetic data designed to mimic real attack scenarios, allowing for faster adaptation and higher detection confidence in live environments.

Fusion Models and Scoring Engine

To ensure that no single indicator is relied on in isolation, Shufti’s detection system fuses over 20 model layers, including video, audio, behavioral, and metadata, into a unified scoring engine. Each model contributes a probability score of fraudulence, and these are combined to create a weighted risk profile.

Our scoring engine supports adaptive thresholds based on the client’s context and regulatory requirements.

For example, a fintech onboarding process might emphasize document authenticity and liveness, while an account recovery process may require stronger behavioral correlations. Clients can configure acceptance, review, or rejection thresholds, model weighting, and override rules (e.g., dark web matches override all other scores). This flexibility enables precise tuning for both false positive reduction and high-risk use cases.

R&D and Continuous Learning

Shufti continuously improves its detection capabilities through ongoing model training and adversarial dataset generation. Because we control our entire tech stack, from data ingestion to model training, we can respond to new and emerging fraud threats in hours, not weeks. Our team regularly analyzes data from both flagged fraud attempts and verified user activity to refine performance, improve accuracy, and reduce false rejections across diverse use cases.

A key part of this process involves the generation of synthetic datasets in-house. We create adversarial samples that blend real and manipulated content to simulate advanced deepfake attacks.

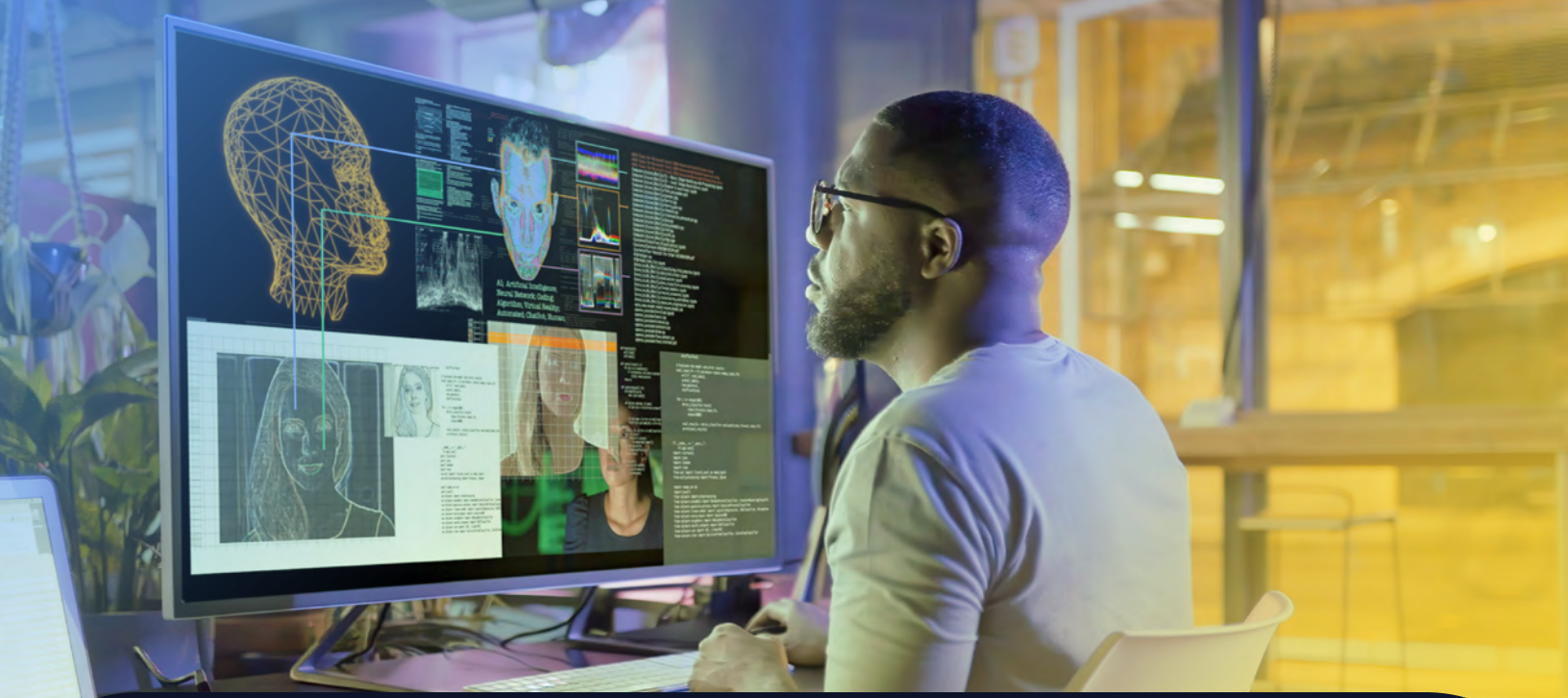
These datasets are used to test our system's limits, uncover blind spots, and train our models to recognize previously unseen forms of synthetic media. This proactive method helps us stay ahead of fraud techniques that are still emerging.

Our models are retrained on an ongoing basis, incorporating insights from real-world activity and synthetic stress testing. This feedback loop ensures that every detection layer benefits from the latest intelligence, allowing our systems to evolve with the threat landscape. As a result, we are able to push detection updates frequently and keep pace with the growing speed and sophistication of deepfake-enabled fraud.

Conclusion and Recommendations

The threat posed by deepfakes has advanced from a fringe concern to a central risk for identity-driven systems across financial services, gaming, insurance, and beyond. As synthetic media becomes more photorealistic, dynamic, and accessible, organizations must recognize that traditional verification methods are designed for an era that has already passed and can no longer offer sufficient protection.

Shufti's multi-layered defense strategy addresses this challenge head-on. Through adaptive liveness checks, behavioral biometrics, multimodal coherence analysis, and continuous model training, our platform provides scalable, real-time fraud resistance designed for the modern digital threat landscape. By owning our entire technology stack, Shufti is able to rapidly deploy new detection layers and stay ahead of the techniques fraudsters use to exploit digital weaknesses. For fraud teams, security leaders, and CTOs, the path forward is clear: proactive adoption of true deepfake-resistant systems is no longer optional. Organizations must invest in solutions that combine layered detection with flexible risk scoring, model transparency, and privacy-by-design principals. Continuous model validation and collaboration across sectors will be critical in reducing exposure to fraud and preventing synthetic identity abuse.



Future-proof against AI-driven fraud

Adaptive deepfake detection must evolve faster than the fraudsters.

Shufti rapidly iterates our approach, with new deepfake detection layers added each month. And when we adapt our models, we seamlessly deploy those changes to our customers — fast.

Protect with a multi-layered defense

Explore our interactive, self-service demo center to see how our solutions can be applied to your organization, on your terms and at your pace.

Explore the Demo Center

WWW.SHUFTI.COM

Shufti