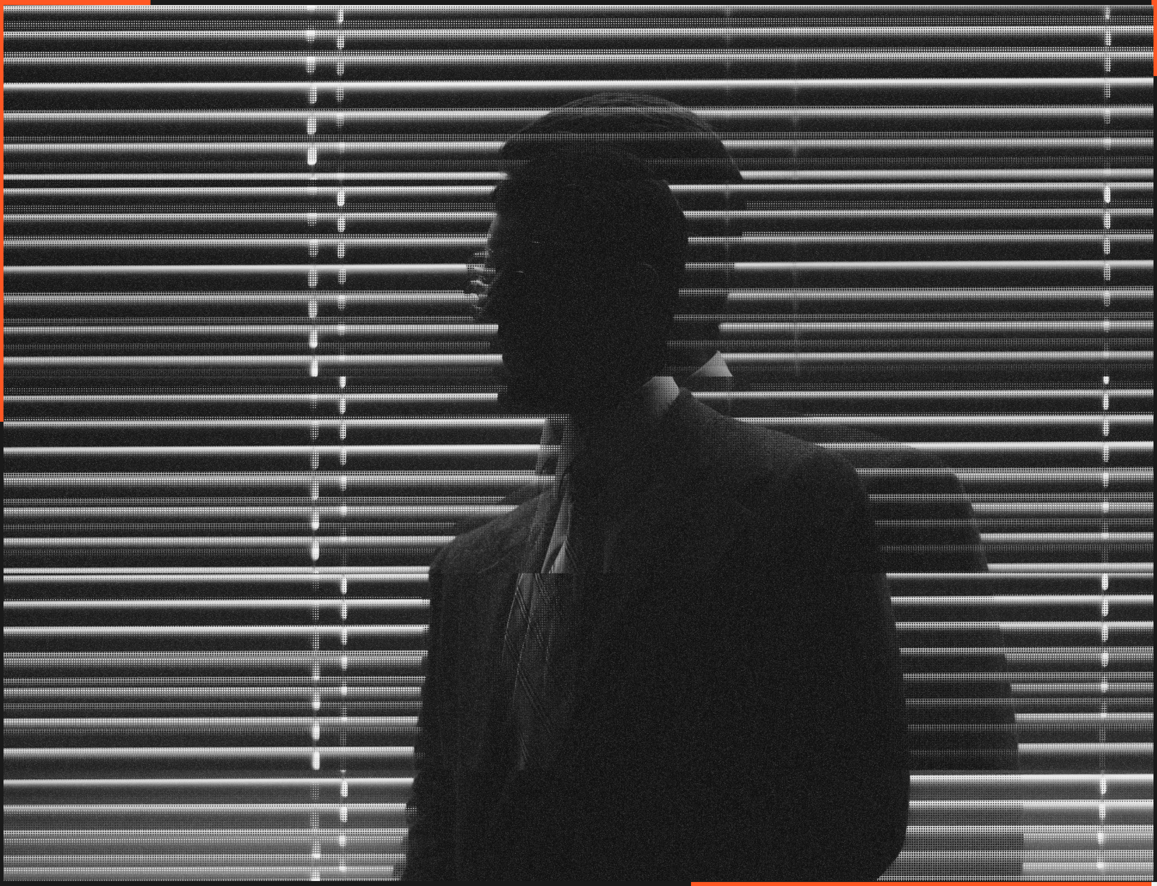


E-book

Deepfakes in Corporate Fraud Campaigns

How Executive Impersonation Has Become a Direct Threat to Corporate Reputation and Security

///AXUR

Executive Summary

The rise of generative AI technologies — especially deepfakes — has unleashed an invisible yet powerful threat: the manipulation of leaders' and authority figures' identities as a weapon for fraud.

Top executives have become strategic targets in campaigns that exploit a new dimension of vulnerability: their public image.

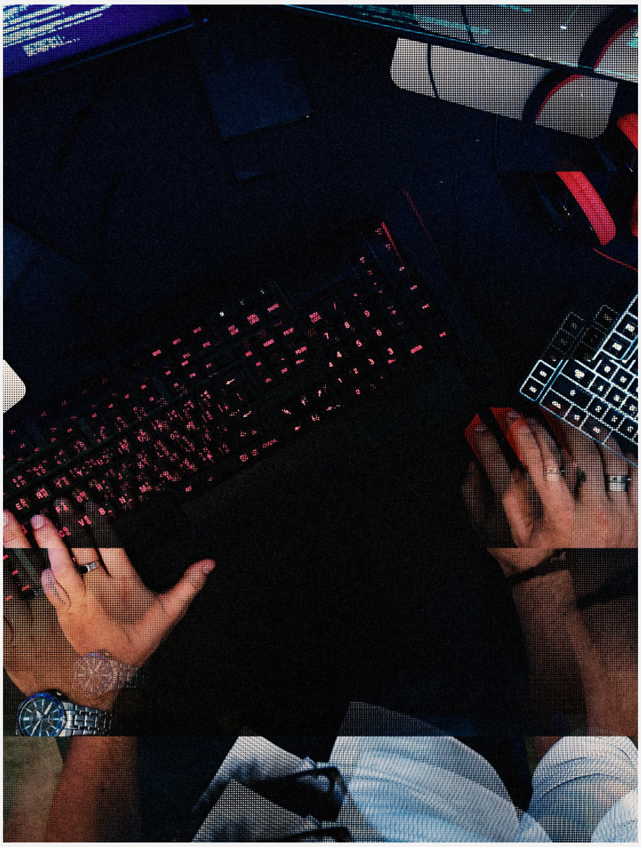
From video calls with cloned voices to fake profiles on social media, attackers are breaking through both technological and psychological barriers to deceive employees, move large sums of money, and damage corporate reputations.

In this eBook, we explore how these campaigns are structured, why executives are squarely in the crosshairs, and how companies can protect themselves. Using our proprietary Fraud Neuron framework, we decode the patterns and tactics behind these operations — and show how AI can be used for both offense and defense.

Deepfakes: The Perfect Illusion in the Service of Fraud

Deepfakes have moved beyond tech curiosities or threats limited to celebrities. Today, they fuel corporate fraud at scale, using the image and voice of executives to pressure employees into critical actions — like authorizing wire transfers, granting system access, or leaking strategic information.

With just a single public image and a few seconds of audio, criminals can create realistic avatars that talk, gesture, and express emotions convincingly. These campaigns are strategic by nature: they rely on social engineering, the authority of executive figures, and urgent scenarios to short-circuit the victim's rational judgment.



Why Executives Are the New Prime Targets

It’s no accident that CFOs, regional directors, and CEOs sit at the center of these frauds. They combine three traits that make them highly valuable to malicious operators:

1. High public exposure: Interviews, corporate videos, online events, and social media offer rich material to train deepfake models.
2. Unquestionable authority: A video or audio command from the "CFO" is rarely challenged by employees — especially under pressure.
3. Access to critical resources: Executives typically hold or influence sensitive financial and operational decisions.

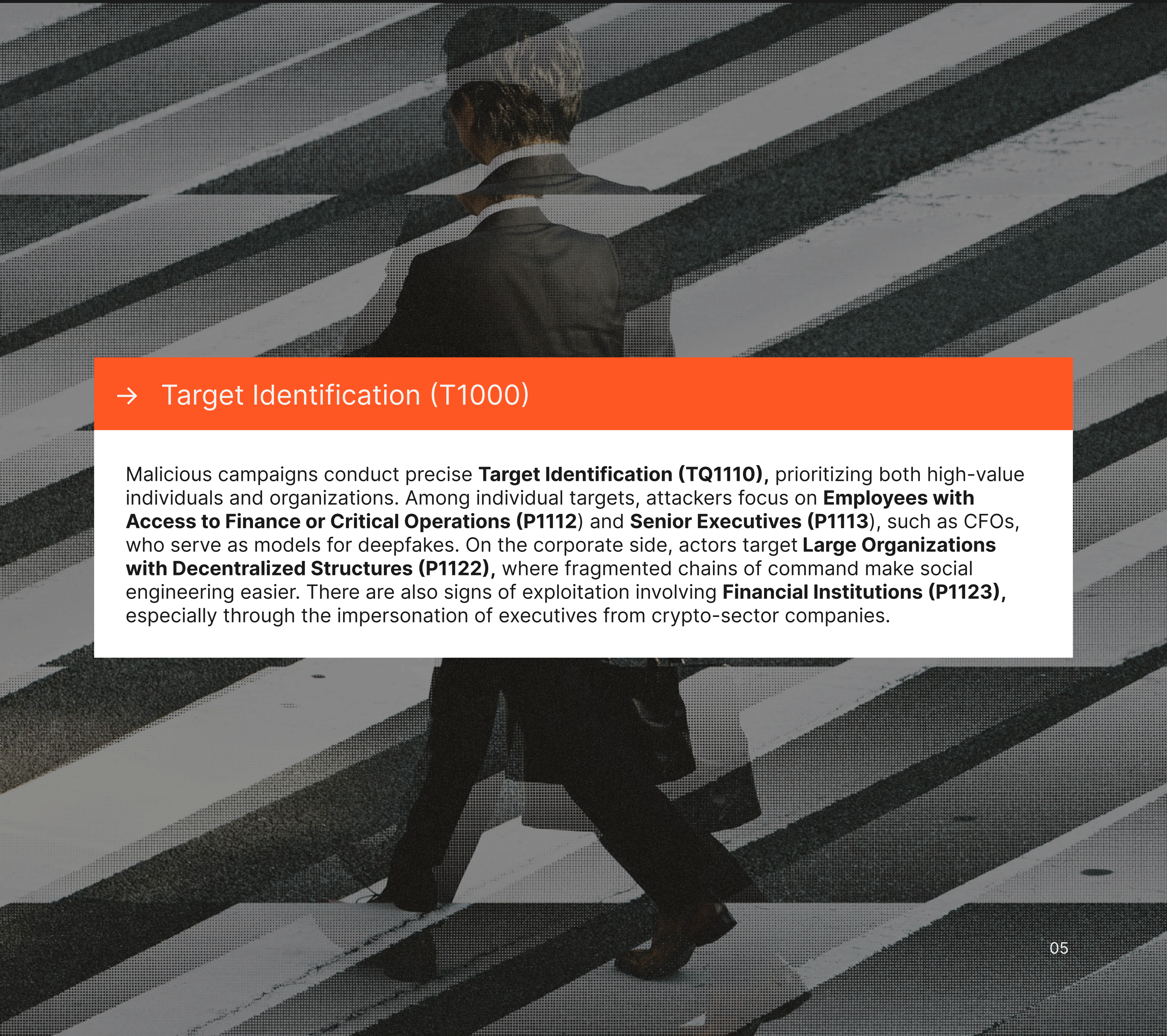
With these variables combined, well-orchestrated campaigns bypass technical controls by exploiting a fundamentally human vulnerability: respect for hierarchy.

Beyond direct financial fraud, there are also indications that these tactics are used for initial access into corporate environments through targeted social engineering.

Anatomy of a Deepfake: Beyond Digital Illusion

To understand the complexity and impact of this phenomenon, we apply the tactical structure from Fraud Neuron — Axur's proprietary framework designed to describe, categorize, and monitor complex digital frauds.

Below, we break down each typical stage of a deepfake-driven campaign, based on real-world incidents.



→ Target Identification (T1000)

Malicious campaigns conduct precise **Target Identification (TQ1110),** prioritizing both high-value individuals and organizations. Among individual targets, attackers focus on **Employees with Access to Finance or Critical Operations (P1112**) and **Senior Executives (P1113**), such as CFOs, who serve as models for deepfakes. On the corporate side, actors target **Large Organizations with Decentralized Structures (P1122),** where fragmented chains of command make social engineering easier. There are also signs of exploitation involving **Financial Institutions (P1123),** especially through the impersonation of executives from crypto-sector companies.

→ Themes (T2000)

The narrative of fraud campaigns is carefully crafted around financial and career opportunity themes. Actors use Investment Offers (P2113) as a pretext — for example, with deepfakes of Binance executives promoting fake crypto opportunities. In parallel, Employment Themes (TQ2200) are exploited through fake job offers (P2211) posted on social media, often featuring deepfake videos of celebrities promoting nonexistent products or services. These themes have high emotional appeal and broaden the reach of fraudulent campaigns.

→ Reconnaissance (T3000)

The reconnaissance phase involves extensive public data collection (TQ3110). Attackers extract visual and voice information from open sources (P3111), such as interviews and public speeches, and mine social media (P3113) to study body language, speech patterns, and organizational context. These datasets feed AI models capable of creating highly convincing deepfakes. There is no evidence of private data leaks or technical infiltration; however, the depth of generated content suggests a meticulous open-source intelligence approach.

→ Resources (T4000)

Operators set up dedicated infrastructure to run their fraud campaigns. They create fake social media accounts (P4113) to impersonate public figures or executives, register deceptive domains (P4211) closely resembling corporate brands, and use server infrastructure (P4212) to host deepfake content or real-time phishing pages. Crypto wallets (P4214) are prepared to receive fraud proceeds. The operation is powered by AI systems (P4311) for video generation and specialized software (P4312) for high-fidelity lip-syncing, voice cloning, and audio manipulation.

→ Conversion (T7000)

The value extraction phase often involves fraudulent wire transfers (P7111) after successfully deceiving victims. In other cases, stolen funds are converted and moved through cryptocurrency transactions (P7113), making them harder to trace.This tactic also applies to scams involving fake promotions and sales, where payments are sent directly to digital wallets controlled by the campaign operators.

→ Identity Simulation (T5000)

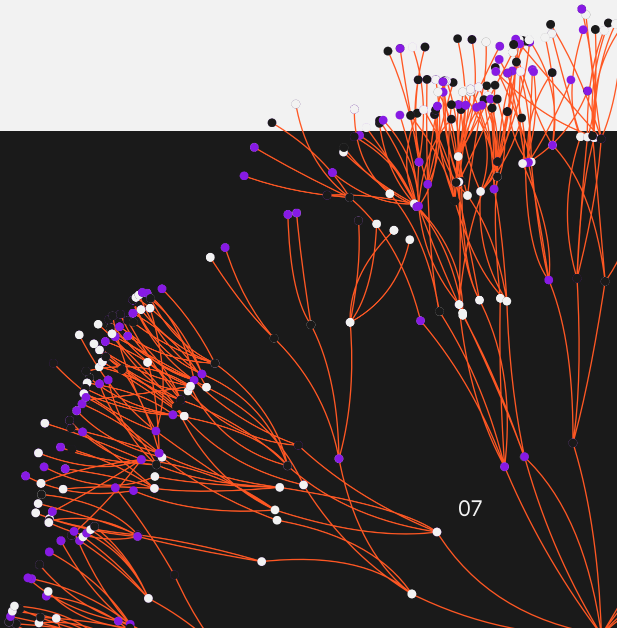
Identity simulation is the core of the campaign. Attackers impersonate employees (P5111), especially senior executives, through video calls and audio messages. They also use deepfakes of public figures (P5113) to create fake promotional videos circulated as paid ads. Tactics include Channel Spoofing (P5213), where attackers simulate Zoom, Microsoft Teams, and other legitimate platforms to create the illusion of authenticity, synchronizing voice and image. The combination of these techniques enables real-time or pre-recorded convincing interactions with victims.

→ Social Engineering (T6000)

The campaign relies heavily on sophisticated social engineering. Operators create Emergency Scenarios (P6112) where a deepfaked executive demands immediate action, usually a wire transfer or access release. They apply Time Pressure (P6121) to reduce victims' critical thinking, forcing snap decisions. By using corporate channels and authoritative imagery, attackers trigger Social Pressure (P6123), making victims feel compelled to obey the "superior's" orders. Psychological manipulation is amplified by the realism of AI-generated voice and image.

→ Impacts (T8000)

Impacts include Direct Financial Losses (P8111), such as the documented case of $25 million stolen through a videoconference deepfake of a CFO. There is also Reputational Damage (P8211) when corporate brands are exploited in fraudulent campaigns, like fake videos featuring Brazilian celebrities. Finally, exposure to these attacks causes Psychological Impact (P8213) on deceived employees who believed they were interacting with real superiors or public figures.



Fraud Neuron

Learn more about Fraud Neuron and help map digital fraud patterns:

github.com/axur/fraudneuron

DeepFake as a Service

The Deepfake-as-a-Service model refers to the direct commercialization of synthetic media technologies powered by artificial intelligence, focused on creating on-demand deepfake videos, audio clips, and avatars.

It follows a typical business structure, where providers offer simplified, low-cost access to complex technologies for clients without advanced technical expertise.

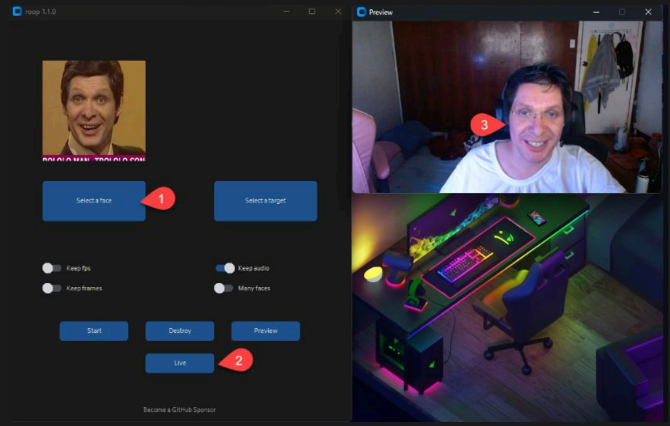
The core feature of this model is the availability of online platforms where buyers can request custom services, such as manipulated video creation, real-time deepfake avatar calls, or voice replication for specific fraud campaigns.

The Deepfake as a Service (DFaaS) market has grown rapidly in recent years, driven by the accessibility and simplicity of generative AI tools. In this business model, commercial platforms deliver advanced technologies that make it easy to create highly realistic manipulated content, even for users with little to no technical experience.

By drastically lowering technical and financial barriers, these tools have not only democratized access to sophisticated technologies but also significantly increased the risk of malicious use of synthetic media.

One notable tool in this space is **Roop**, an open-source application that gained popularity by allowing users to swap a person's face into any video using just a single reference image — with no need for model training or large datasets.

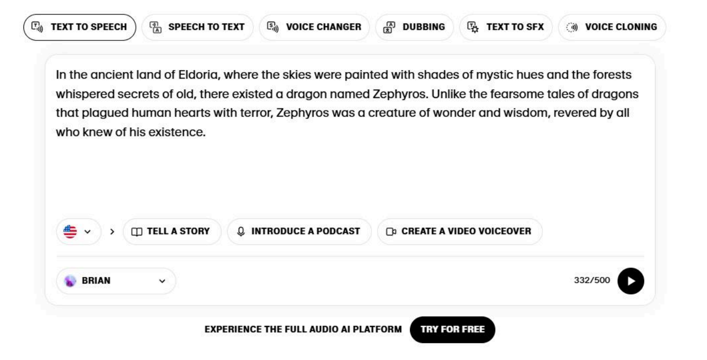
Roop uses pre-trained neural networks to perform near-instant face swaps, producing convincing visual results even with minimal technical investment (Roop, GitHub, 2023).



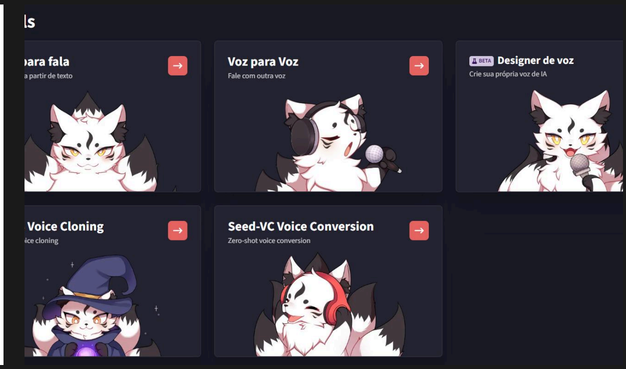
In addition, there are platforms offering advanced voice cloning and speech synthesis technologies, such as ElevenLabs, which can also be exploited for malicious purposes.

With this technology, users can generate highly realistic audio from short original voice samples, replicating emotional nuance, rhythm, and intonation with extreme fidelity.

The widespread availability of this tool has been cited as a decisive factor in successful social engineering attacks, including recently documented cases involving executives at major corporations, where cloned voices were used to carry out significant financial fraud (CNN, 2024; Fortune, 2024).



In the ancient land of Eldoria, where the skies were painted with shades of mystic hues and the forests whispered secrets of old, there existed a dragon named Zephyros. Unlike the fearsome tales of dragons that plagued human hearts with terror, Zephyros was a creature of wonder and wisdom, revered by all who knew of his existence.



A similar tool, known as FakeYou, provides simplified voice synthesis based on typed text.

FakeYou stands out for offering a large catalog of pre-trained synthetic voices, including fictional characters and celebrities, making the creation of synthetic content even easier.

The platform's simplicity has made it popular not only for legitimate uses but also as a quick tool for fraudulent campaigns and disinformation efforts (FakeYou, 2023).

Our new project: TalkingAvatar.ai: a perfect AI Avatar tool by voice clone and lip-syncing!

Create face swap live streams in a second!

The tool known as Nudefusion uses artificial intelligence to transform ordinary photos into fake explicit images — a process commonly referred to as "deepnude." This service raises serious concerns about potential abuse, privacy violations, and defamation. The tool is openly promoted, offering quick results from a simple image upload, with no prior training or technical skills required from the user.

Due to its high potential for misuse, similar platforms have already been subject to investigations and legal actions in several countries, clearly demonstrating the social and ethical impact of deepfake technologies when applied to the non-consensual creation of fake intimate images..

High-Profile Attacks: Real Cases and Techniques Used

Scenario #1 Corporate Sabotage

In this scenario, criminals use deepfakes to spread disinformation, damage brand reputations, manipulate market perception, or disrupt merger and acquisition (M&A) negotiations. Sabotage is carried out by creating fake videos and audios showing executives or employees in compromising situations or spreading false information about products and corporate strategies.

Real Case

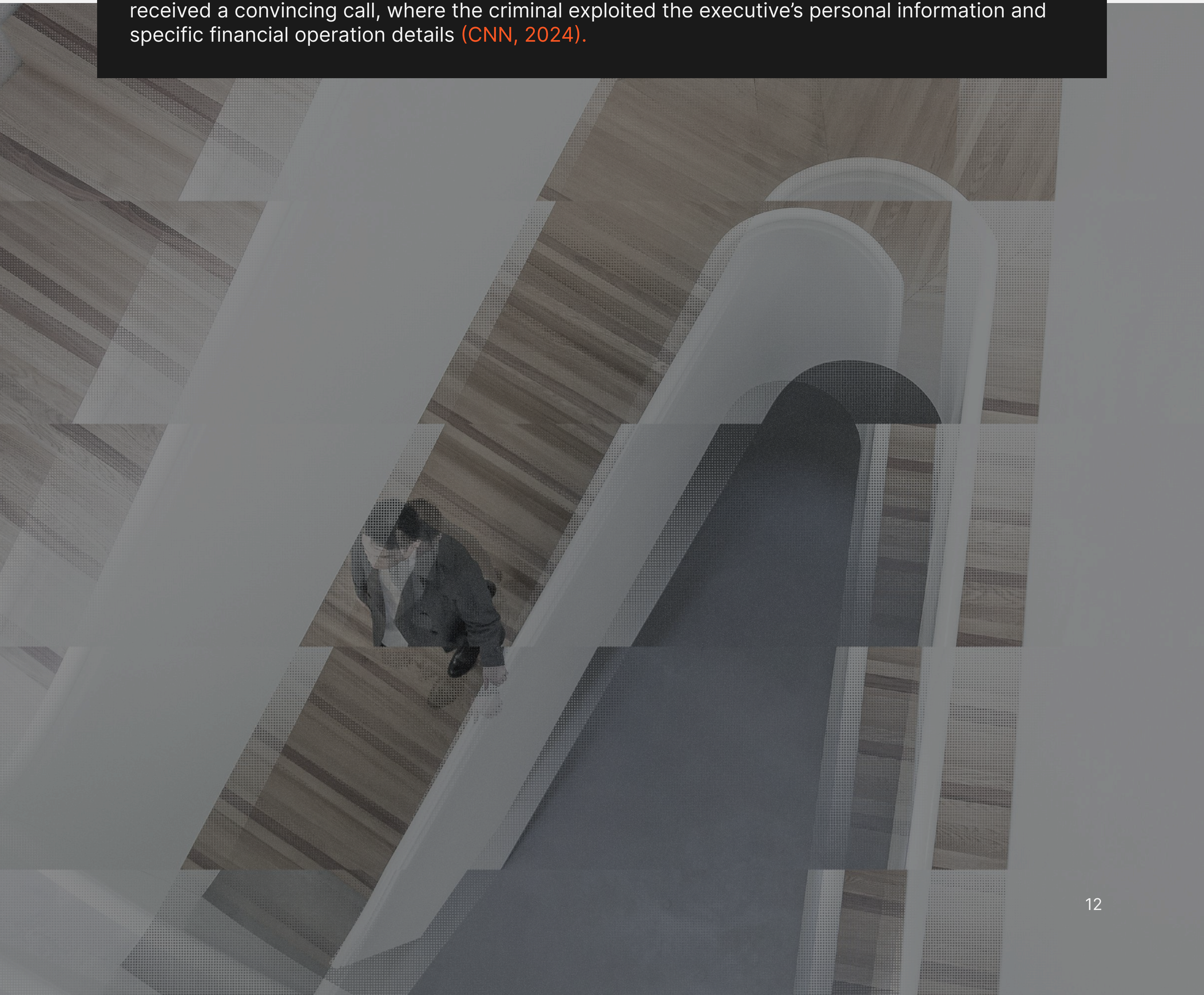
A recent example is the attack against Binance, where criminals used a deepfake hologram of executive Patrick Hillman to manipulate information about fake financial projects, damaging the credibility of the exchange (Bitdefender, 2023).

Scenario #2 Advanced Social Engineering in the Corporate Environment

In this scenario, criminals use deepfakes to carry out highly convincing social engineering attacks targeting senior executives or strategic employees. The attacker first conducts detailed research on executives, gathering personal data from social media and audiovisual materials to train advanced models. After collecting information about recent corporate events — such as new joint ventures or mergers — the attacker creates audiovisual deepfakes of the CEO or CFO, using cloned audio to perform fraudulent calls.

Technical Example and Real Case

In February 2024, criminals executed a sophisticated attack in Hong Kong using a deepfake of a major corporation's CFO. A fraudulent transfer of over $25 million was authorized after the victim received a convincing call, where the criminal exploited the executive's personal information and specific financial operation details (CNN, 2024).



Scenario #3 Social Engineering Attacks Against Financial Institutions

In this scenario, attackers use deepfake audio to compromise voice authentication systems at banks and financial institutions. After purchasing stolen personal data from the dark web (such as names, national IDs, and bank account numbers) and collecting audiovisual samples of victims from social media, criminals create voice clones to fool authentication systems and gain access to financial accounts.

Technical Example and Real Case

While no detailed public records exist for this exact scenario, similar techniques have been identified in attacks against financial institutions, where voice cloning was used to deceive call center employees, enabling fraudulent transfers and unauthorized access to accounts — demonstrating clear technical feasibility (Fortune, 2024; BleepingComputer, 2024).

Scenario #4 Corporate Liability and Deepfake Fraud

In this scenario, criminals create realistic deepfake videos suggesting major failures or accidents involving corporate products or services, aiming to fraudulently obtain financial compensation.The attacker conducts detailed research on past real incidents and crafts a false visual narrative, using advanced face-swapping models to simulate fictitious victims.

Scenario #5 Market Manipulation and Stock Price Fraud via Deepfake

In this scenario, attackers use deepfakes to manipulate stock prices in financial markets. The criminal first acquires a position in a stock, then creates and distributes fake content on popular forums like Reddit or Stockaholics. Convincing deepfakes of CEOs or executives announcing fake financial events prompt investors to quickly buy or sell, causing significant fluctuations in the asset's price.

Technical Example and Real Case

While deepfake-specific cases targeting stock market manipulation have not been widely reported so far, similar scenarios have been observed in cryptocurrency frauds — such as the fake giveaway involving Solana's founder, Anatoly Yakovenko. In that incident, criminals used facial deepfakes during fake YouTube live streams to drive fraudulent transactions (Reddit, 2024).

How to Detect and Mitigate Deepfakes: A Technical Guide for Companies and Individuals

Can We Educate the Public to Spot Deepfakes?

Public education and awareness are crucial to effectively combat the spread of synthetic media.

However, research suggests that public engagement in this learning process may be limited.

Studies by PAI show that people prefer not to be instructed directly or in a patronizing way about what is true or false. Instead, they prefer learning through tools and contexts that allow them to independently verify media authenticity.

Still, it is possible to empower individuals by teaching them to spot clear technical signals, such as:

Signs that a Video or Image Might Be a Deepfake:

→ Blurred or smudged areas around the face, inconsistent with the rest of the image.

→ Sudden changes in skin tone or lighting around the face.

→ Duplicated edges (e.g., double chin, duplicated eyebrows).

→ Lack of blinking, excessive blinking, or unnatural facial movements.

→ Lip-sync issues or unnatural mouth movements.

→ Inconsistencies between the background and the main subject.

Signs that an Audio Might Be a Deepfake:

→ Choppy phrasing or unnatural speech rhythms.

→ Abrupt changes in voice tone.

→ Unusual word choices or phrases the real speaker would not typically use.

→ Context that feels off-topic or an inability to answer simple, related questions.

Signs that a Text Might Be Manipulated or Fraudulent:

→ Unusual spelling or grammatical errors.

→ Lack of logical coherence and flow.

→ Unknown or suspicious email addresses or phone numbers.

→ Uncharacteristic expressions or vocabulary for the supposed sender.

→ Out-of-context or unexpected messages.

However, as technology advances, distinguishing manipulated media from authentic content is becoming increasingly complex.

Deepfakes are evolving rapidly, driven by constant improvements in generative models and sophisticated AI techniques (such as GANs and autoencoders).

As a result, relying solely on human perception to detect fraudulent content is no longer sufficient.

To increase detection effectiveness, it is necessary to integrate robust technical approaches, including digital forensic analysis and automated algorithms.

How Can We Improve Detection Capabilities?

Today, tools like Reality Defender, Microsoft Video Authenticator, Deepware AI, and academic frameworks such as FaceForensics++ are available to analyze deepfakes with increasing precision. These systems apply detailed analysis using machine learning algorithms specifically trained to detect unique features, such as lighting inconsistencies, unnatural facial expressions, or digital artifacts left behind during the manipulation process.

However, no tool is entirely foolproof — especially against highly sophisticated and technically advanced deepfakes. Detection effectiveness depends directly on continuous model training, constant fine-tuning, and regular exposure to new manipulated examples.

Moreover, the tools mentioned do not provide advanced monitoring to detect fraud involving executives' images across social media, websites, and streaming platforms — which is critical for gaining visibility into these threats.

How Are These Threats Connected to Brand Protection and Fraud?

Deepfake campaigns involving executives and public figures don't just impact direct victims. When they occur, the effects ripple across the entire corporate structure — damaging institutional reputation, affecting stakeholder relationships, and often resulting in financial losses. The misuse of organizational leaders' identities isn't just a personal violation; it's an entry point for large-scale corporate fraud.

Use of Executive Identities in Fake Promotions and Deceptive Campaigns

In this type of scam, a well-known executive appears in a video or audio clip promoting investments, announcing partnerships, or offering advantages — none of which actually exist. In many cases, this fake content is sponsored and distributed on social media as if it were part of the company's official communications. The direct consequence is brand damage. Employees, customers, and partners may associate the fake content with the organization's legitimate communications, causing confusion and eroding trust.

Even after containing the incident, rebuilding credibility can be a long and costly process.

It's important to note that criminals don't necessarily need to use the victim's face—distinctive physical features or signature style elements can also prompt mistaken associations.

Linking Targeted Attacks to Brand Incidents

Another critical point is the overlap between social engineering attacks and brand abuse. Criminals are increasingly combining executive deepfakes with fake domains, visually legitimate emails, and cloned social media profiles to create believable scenarios.

The use of corporate identity elements in these attacks — logos, brand names, slogans, and even aspects of institutional tone of voice — further strengthens the illusion of authenticity. For information security teams, this presents a challenge that goes beyond detecting technical threats.

It requires monitoring signs of brand and image abuse across multiple environments and responding quickly to prevent isolated incidents from escalating into full-blown institutional crises.

Responding to this type of attack demands more than monitoring; it requires direct intervention in the environments where fraudulent content is being distributed. This is where the concept of takedown comes into play — a process that involves requesting the removal of malicious content or assets from platforms, service providers, social networks, domain registrars, and other intermediaries.

What Can Be Taken Down?

In cases of deepfake-related fraud, takedown becomes an essential tool when the manipulated content violates the usage guidelines of the platform where it is hosted. Fake executive profiles on social media, videos hosted on third-party websites, deceptive ads, and phishing pages using corporate branding — all of these elements are generally eligible for removal.

The U.S. federal Take It Down Act requires platforms primarily dedicated to user-generated content to provide a dedicated and fast-track channel for reporting and removing intimate visual depictions—whether authentic or AI-generated, such as deepfakes—within 48 hours. The law does not require the image to be pornographic or to show the person's face; any identifiable feature, such as a birthmark, may be sufficient grounds to request a takedown under this legislation.

However, not all AI-generated **content can be taken down easily**. Content that portrays fictional scenarios without directly infringing on a trademark, for example, may fall outside traditional moderation criteria. In such cases, the role of information security shifts to assessing reputational risk, gathering technical evidence of manipulation, and activating alternative response channels — such as corporate communications, legal action, and crisis management.

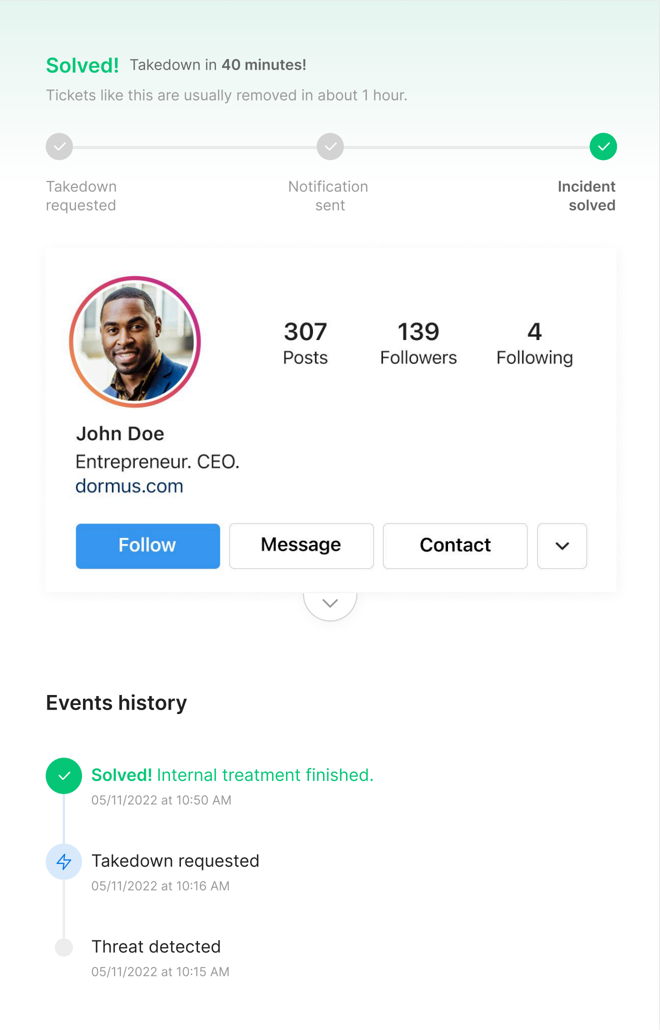
What Can (or Cannot) Be Taken Down in a Deepfake Fraud Case

Fraud Element Used	Domain	Takedown Possible?
Email impersonating an executive using a domain similar to the official one	Corporate	✓ **Yes** The domain can be removed
Fake profile using the executive's name and photo on LinkedIn or Instagram	Personal (internal impact)	✓ **Yes** The profile can be removed
Fake page with the company's branding (logos, colors, trade name)	Corporate	✓ **Yes** It can be removed for trademark violation
Deepfake video of an executive promoting a fake investment using the brand	Both	✓ **Yes** Removable from platforms and ads
Image or video showing an intimate scene shared without consent, whether authentic or fake	Both	✓ **Yes** Under the Take It Down Act, it must be removed within 48 hours starting in 2026
Deepfake video of an executive without brand use but realistic appearance	Personal	✓ **Yes** Removable
Video on a foreign platform simulating a generic executive, without brand use	Both	× **No** Difficult to remove without explicit violation
Fake domain simulating an official subdomain (e.g., finance.yourcompany.online)	Corporate	✓ **Yes** Removable based on brand abuse
Cloned audio of an executive sent via messaging apps	Personal (internal impact)	⚠ **Not directly** Difficult to track or remove
Fake video shared in private groups (WhatsApp, Telegram, etc.)	Both	⚠ **Not directly** But channels can be monitored and investigated
Fake post simulating an official statement signed by an executive	Corporate	✓ **Yes** Removable for image and brand violation

The Axur platform enables large-scale identification and takedown requests, supporting different types of assets and channels.

In scenarios where response speed is critical — such as a fraud campaign using the CEO's image circulating through a fake profile — reaction time can be the decisive factor between containing an incident and allowing it to escalate.

For security teams, this capability means not only faster response but also a stronger integration point with other areas of the company, expanding visibility and coordination during high-impact external incidents.



Solved! Takedown in **40 minutes!**

Tickets like this are usually removed in about 1 hour.

✓ Takedown requested — ✓ Notification sent — ✓ **Incident solved**

307	139	4
Posts	Followers	Following

John Doe
Entrepreneur. CEO.
dormus.com

Events history

✓ **Solved!** Internal treatment finished.
05/11/2022 at 10:50 AM

Takedown requested
05/11/2022 at 10:16 AM

Threat detected
05/11/2022 at 10:15 AM

Conclusion

As we've seen throughout this material, the attack surface is no longer limited to systems. It now extends to the institutional identity and public presence of organizational leaders — which, when manipulated, can become tools for scams with the potential to cause financial losses, reputational crises, and strategic impacts.

The trend is that these attacks will become even more convincing. Real-time voice cloning technologies, interactive avatars, and the integration of deepfakes with large language models (LLMs) are already making it possible to create synthetic executives capable of personalized, adaptive interactions with victims.

The scalability of these attacks will increasingly demand a coordinated response between security, legal, communications, and governance teams.

The Axur platform was developed to support companies in this new landscape. We automate the detection of external threats, the monitoring of fake profiles, and the execution of takedowns at scale, offering visibility and agility to contain fraud involving executive identities and brand reputation. If your company is already dealing with incidents of this kind — or wants to prepare to face them with confidence — we are ready to support your defense strategy.

Discover Our Solution for Executives and VIPs

DISCOVER MORE

bsi ISO/IEC 27001 Information Security Management CERTIFIED

CleanDNS Trusted Reporter

///AXUR