RESEMBLE.AI

Deepfake Threats

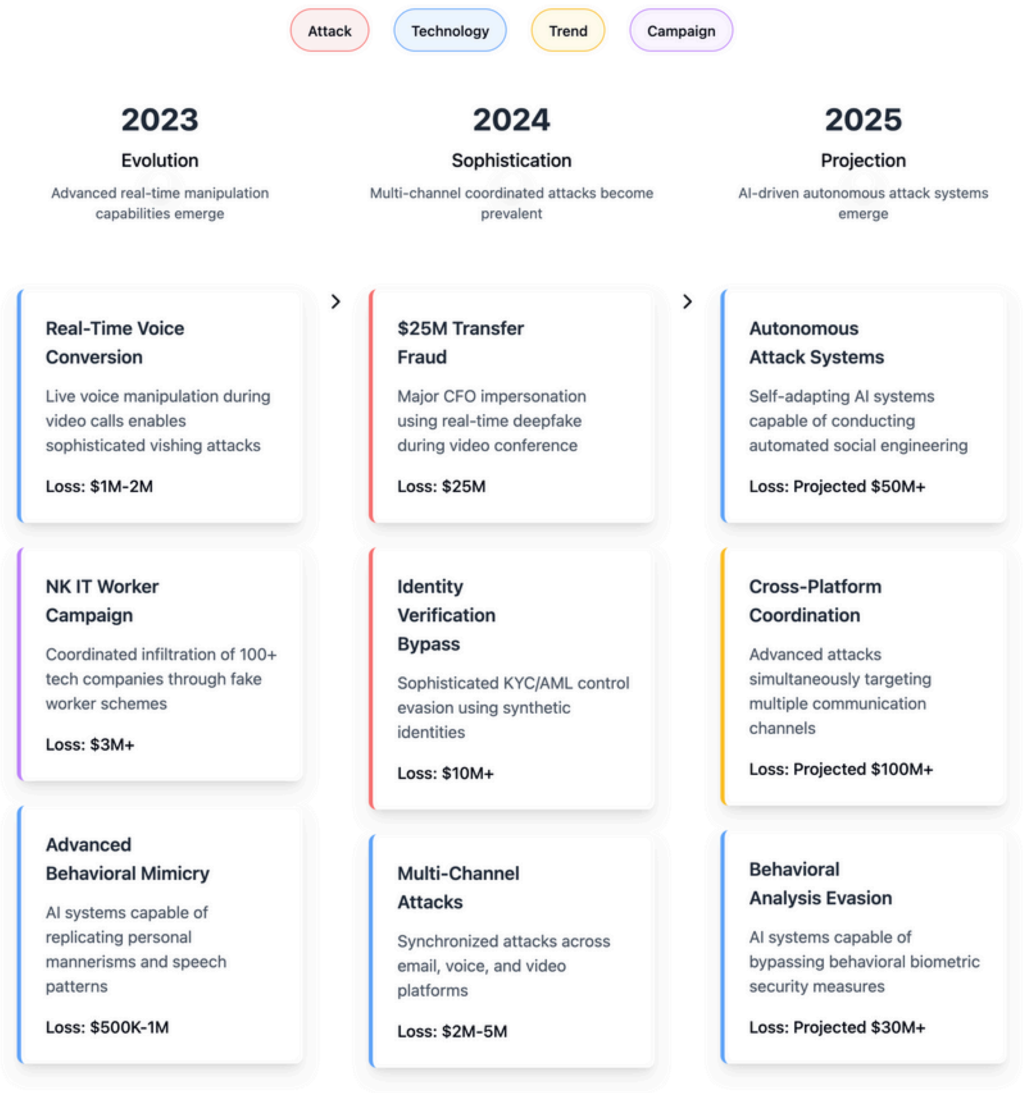
An Analysis of Enterprise Risk and Mitigation in 2024

Executive Summary

The synthetic media landscape has fundamentally transformed in 2024. A recent incident where a finance employee authorized a $25 million transfer during what appeared to be a legitimate video conference with their CFO exemplifies this evolution. This is not an isolated case - enterprises across sectors report sophisticated deepfake attacks targeting their core operations, from executive communications to recruitment processes.

Technical Evolution of Deepfakes

The synthetic media landscape has fundamentally transformed in 2024. A recent incident where a finance employee authorized a $25 million transfer during what appeared to be a legitimate video conference with their CFO exemplifies this evolution. This is not an isolated case - enterprises across sectors report sophisticated deepfake attacks targeting their core operations, from executive communications to recruitment processes.



Attack | Technology | Trend | Campaign

2023
Evolution
Advanced real-time manipulation capabilities emerge

Real-Time Voice Conversion
Live voice manipulation during video calls enables sophisticated vishing attacks
Loss: $1M-2M

NK IT Worker Campaign
Coordinated infiltration of 100+ tech companies through fake worker schemes
Loss: $3M+

Advanced Behavioral Mimicry
AI systems capable of replicating personal mannerisms and speech patterns
Loss: $500K-1M

2024
Sophistication
Multi-channel coordinated attacks become prevalent

$25M Transfer Fraud
Major CFO impersonation using real-time deepfake during video conference
Loss: $25M

Identity Verification Bypass
Sophisticated KYC/AML control evasion using synthetic identities
Loss: $10M+

Multi-Channel Attacks
Synchronized attacks across email, voice, and video platforms
Loss: $2M-5M

2025
Projection
AI-driven autonomous attack systems emerge

Autonomous Attack Systems
Self-adapting AI systems capable of conducting automated social engineering
Loss: Projected $50M+

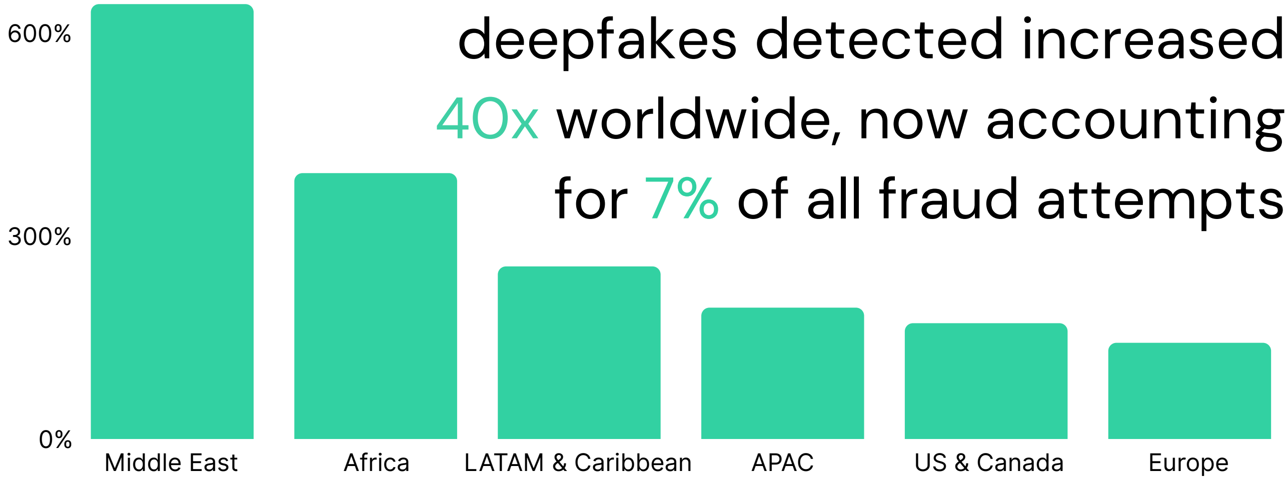
Cross-Platform Coordination
Advanced attacks simultaneously targeting multiple communication channels
Loss: Projected $100M+

Behavioral Analysis Evasion
AI systems capable of bypassing behavioral biometric security measures
Loss: Projected $30M+

RESEMBLE.AI

Enterprise Infiltration: The North Korean Example

A particularly concerning development in 2024 involves the sophisticated use of deepfake technology in corporate infiltration attempts. In a recent case documented by Exabeam, North Korean operatives attempted to penetrate the company's security team using deepfake technology during video interviews. During the interview process for a senior governance, risk, and compliance analyst position, the candidate displayed sufficient technical knowledge to pass initial HR screenings. However, as Exabeam's CISO Kevin Kirkwood noted, subsequent technical interviews revealed telltale signs of synthetic media - mechanical responses, poor lip synchronization, and unnatural eye movements.



From 2022 to 2024, number of deepfakes detected increased 40x worldwide, now accounting for 7% of all fraud attempts

2024 DEEPFAKE FRAUD GROWTH BY REGION

This incident is part of a broader pattern. Over 300 businesses have fallen victim to such fake worker IT scams, with individual North Korean IT workers earning upwards of $300,000 annually through these deceptions. According to a joint advisory from the U.S. Department of State, Treasury, and FBI, teams of these operators can collectively generate more than $3 million annually.

Authentication and Detection Challenges

Traditional security controls are proving inadequate against modern synthetic media attacks. According to NIST's 2024 analysis, enterprises face a fundamental challenge: their authentication systems were designed for a threat landscape where content manipulation was static and detectable through forensic analysis. Traditional content authentication methods - digital signatures, watermarking, and metadata verification - are proving inadequate against modern synthetic media attacks.

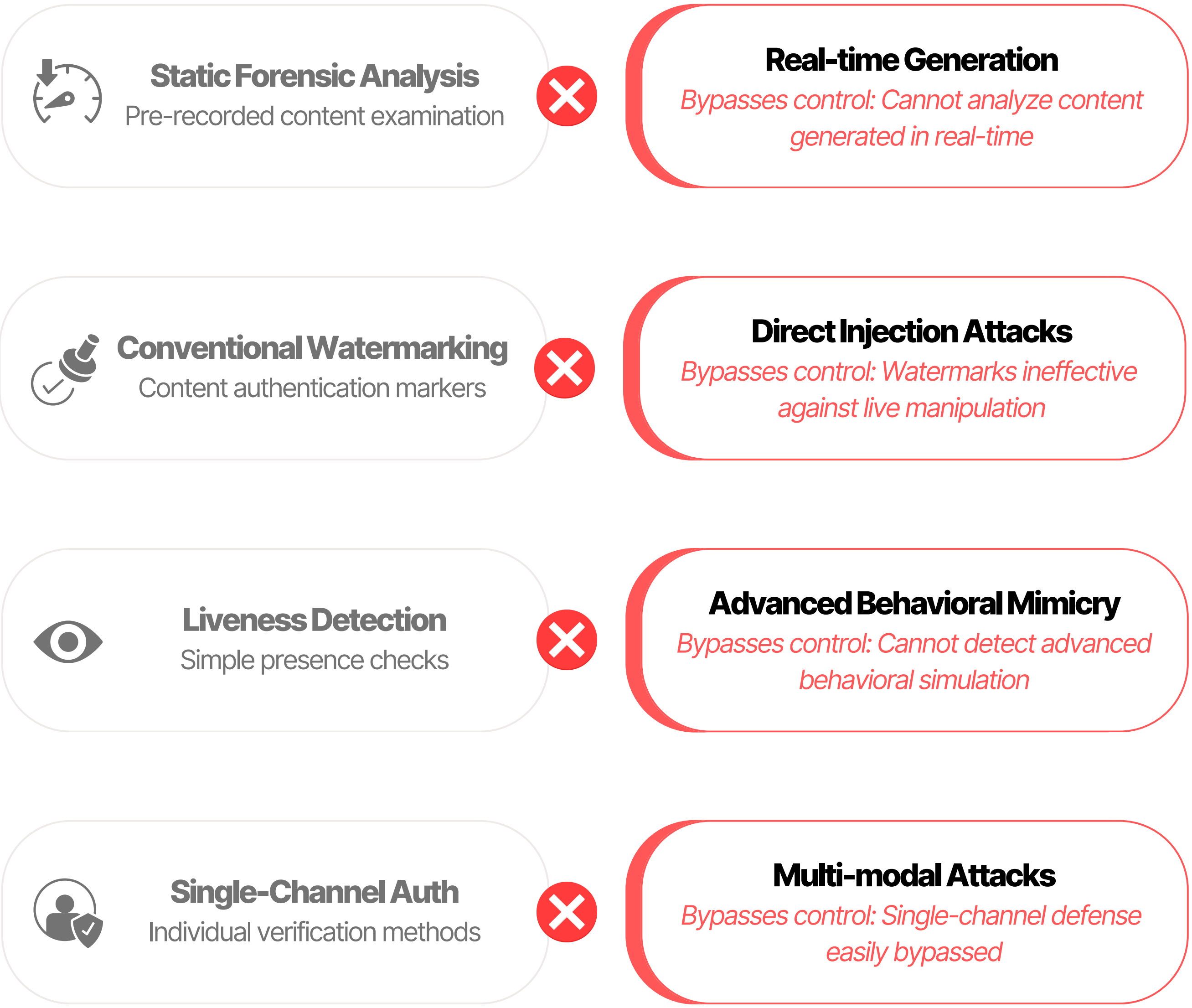
Authentication Challenges: Current Security Gaps

Traditional security controls (digital signatures, watermarking, metadata verification) are inadequate against modern synthetic media attacks that combine multiple deception techniques and operate in real-time.

Today's synthetic media attacks operate differently, attackers combine multiple deception techniques, from document forgery to real-time video manipulation. Content is synthetically generated during live interactions, making traditional forensic analysis ineffective. Malicious content is injected directly into verification processes, bypassing standard security controls. Advanced deepfakes can mimic natural human behavior, defeating conventional liveness detection. Multi-modal attacks combining video, audio, and behavioral manipulation present complex detection challenges that exceed traditional security capabilities.

Security Control Inadequacies

Why Traditional Controls Fail Against Modern Synthetic Threats



Static Forensic Analysis
Pre-recorded content examination

✗

Real-time Generation
Bypasses control: Cannot analyze content generated in real-time

Conventional Watermarking
Content authentication markers

✗

Direct Injection Attacks
Bypasses control: Watermarks ineffective against live manipulation

Liveness Detection
Simple presence checks

✗

Advanced Behavioral Mimicry
Bypasses control: Cannot detect advanced behavioral simulation

Single-Channel Auth
Individual verification methods

✗

Multi-modal Attacks
Bypasses control: Single-channel defense easily bypassed

RESEMBLE.AI

Industry-Specific Impacts

The technology sector has emerged as a primary target for sophisticated synthetic media attacks. CrowdStrike documented that one North Korean group (Famous Chollima) successfully infiltrated over 100 companies through impersonation campaigns. These attacks specifically targeted companies with valuable intellectual property and strategic technologies, demonstrating the strategic nature of synthetic media threats.

The financial services sector faces unique challenges with synthetic media authentication. FinCEN's 2024 alert highlights how attackers are successfully opening accounts using AI-generated identities to facilitate various fraud schemes. These synthetic identities are increasingly sophisticated, combining deepfake images with stolen or fabricated personal information in ways that defeat traditional Know Your Customer (KYC) controls.

Manufacturing and critical infrastructure sectors face particularly concerning threats. The NSA, FBI, and CISA joint cybersecurity advisory emphasizes how synthetic media attacks against these sectors could have cascading effects across supply chains and essential services. Attackers can use deepfake technology to compromise industrial control systems by deceiving operators or bypassing physical security measures through synthetic credential generation.

RESEMBLE AI

Major Deepfake Incidents Timeline

2022

March 16

Political Disinformation

Zelenskyy Surrender Deepfake

A deepfake video depicted Ukrainian President Volodymyr Zelenskyy urging his soldiers to surrender during the Russian invasion. The video was briefly broadcast on Ukrainian television and circulated on social media before being debunked.

Dec 22

Propaganda

"Wolf News" Synthetic Anchors

Pro-China propagandists disseminated deepfake videos featuring synthetic news anchors in a fictitious "Wolf News" segment, promoting Chinese government narratives.

2023

March 28

Misinformation

Pope Francis Balenciaga

An AI-generated image of Pope Francis wearing a white Balenciaga puffer jacket went viral, misleading many into believing it was real.

Oct 09

Political Misinformation

Keir Starmer Audio

A deepfake audio clip falsely portrayed UK Labour Party leader Keir Starmer verbally abusing staffers, coinciding with the party's conference.

RESEMBLE AI

Major Deepfake Incidents

Timeline

2023

Oct 10

Political Misinformation

Slovak Election Interference

An audio deepfake of Slovak politician Michal Šimečka was released, falsely suggesting he discussed rigging the upcoming election.

2024

Feb 08

Election Interference

Biden Robocall Incident

Over 20,000 New Hampshire voters received robocalls featuring an AI-generated voice of President Joe Biden, urging them not to vote in the Democratic primary. This led to legal actions against the perpetrators.

April 25

Political Disinformation

Marcos Military Action Hoax

A deepfake audio clip misrepresented Philippine President Bongbong Marcos, suggesting he ordered military action in response to tensions in the South China Sea.

July 22

Defamation

Marcos Defamation Video

A deepfake video surfaced, falsely depicting President Marcos engaging in illicit activities. Investigations confirmed the video's AI-generated nature.

RESEMBLE AI

Major Deepfake Incidents

Timeline

2024

Sept 03

Election Interference

Spamouflage Campaign

The "Spamouflage" network, linked to Chinese actors, employed deepfake videos to impersonate American voters, spreading divisive narratives ahead of the U.S. presidential election.

May 15

Financial Fraud

Queensland Premier Scam

Scammers used AI to clone the voice of Queensland Premier Steven Miles, promoting a fraudulent investment scheme. The deepfake was convincing enough to deceive potential investors.

May 26

Political Misinformation

Sunak Military Service Hoax

A deepfake audio clip falsely depicted UK Prime Minister Rishi Sunak announcing mandatory military service for 18-year-olds in conflict zones, causing public concern before being debunked.

Vectors of attack

As deepfake technology continues to advance, enterprises must be aware of the various attack vectors and high-risk areas that malicious actors may target. Some of the most vulnerable points of entry and exploitation include:

1. Executive Communications: Deepfake audio and video can be used to impersonate C-suite executives, tricking employees into disclosing sensitive information, transferring funds, or granting unauthorized access to critical systems. Fraudulent video calls, emails, and voice messages purporting to be from senior leadership are becoming increasingly difficult to detect.

2. Financial Transactions: Synthetic identities created using deepfake technology can be used to open fraudulent accounts, conduct money laundering, or initiate unauthorized financial transactions. The financial services sector is particularly vulnerable to these attacks, as traditional identity verification methods struggle to detect sophisticated AI-generated identities.

!

Attack Vectors: Key Vulnerability Points

Executive communications, financial transactions, HR/recruitment, social engineering, IP theft, supply chain compromise, and reputational damage through fake media.

Vectors of attack

3. Human Resources and Recruitment: As demonstrated by the North Korean IT worker case studies, deepfakes can be used to create synthetic identities and fake credentials to infiltrate organizations through recruitment processes. Once inside, these malicious actors can steal sensitive data, plant malware, or conduct espionage.

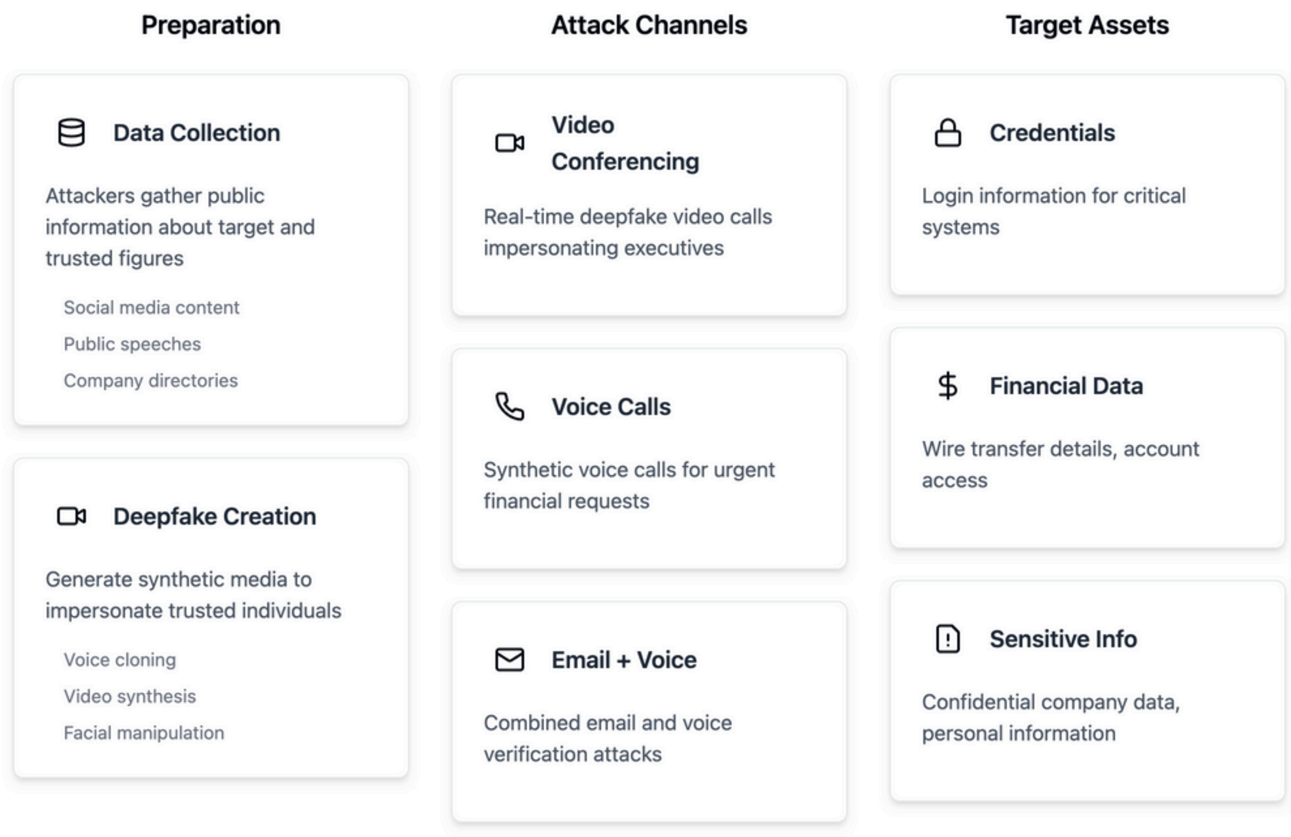
4. Social Engineering and Phishing: Deepfake audio and video can make social engineering attacks, such as phishing or vishing (voice phishing), much more convincing. By impersonating trusted individuals, attackers can manipulate employees into revealing login credentials, financial information, or other sensitive data.

5. Intellectual Property Theft: Nation-state actors and competitors may use deepfake technology to infiltrate organizations and steal valuable intellectual property, such as trade secrets, research and development data, or confidential business strategies. The technology sector is a prime target for these attacks.

Vectors of attack

6. Supply Chain Compromise: Deepfakes can be used to impersonate suppliers, vendors, or business partners, tricking companies into sharing sensitive information or allowing unauthorized access to their networks. This can lead to supply chain disruptions, product tampering, or the installation of malicious software.

7. Reputational Damage: Deepfake technology can be used to create fake news, misleading social media posts, or manipulated videos that damage a company's reputation. These attacks can cause significant financial losses, erode customer trust, and even impact stock prices.



Preparation

Data Collection

Attackers gather public information about target and trusted figures

- Social media content
- Public speeches
- Company directories

Deepfake Creation

Generate synthetic media to impersonate trusted individuals

- Voice cloning
- Video synthesis
- Facial manipulation

Attack Channels

Video Conferencing

Real-time deepfake video calls impersonating executives

Voice Calls

Synthetic voice calls for urgent financial requests

Email + Voice

Combined email and voice verification attacks

Target Assets

Credentials

Login information for critical systems

Financial Data

Wire transfer details, account access

Sensitive Info

Confidential company data, personal information

Emerging Defense Initiatives

Regulatory Compliance and Risk Management

The regulatory landscape around synthetic media continues to evolve. The California AI Transparency Act represents a significant step forward, establishing specific requirements for synthetic content detection and disclosure. Organizations must now prepare for similar regulations in other jurisdictions, as government agencies worldwide recognize the growing threat posed by deepfake technologies.

Risk management strategies must evolve accordingly. The NSA, FBI, and CISA joint advisory emphasizes the importance of treating synthetic media threats as a core component of enterprise risk management. Organizations must develop comprehensive risk assessment frameworks that consider both direct financial impacts and broader operational consequences of synthetic media attacks.

SEPTEMBER 2023

DEEPFAKES ACCOUNTABILITY ACT

Bill introduced by Rep. Yvette Clarke to establish transparency requirements and penalties for malicious deepfake use

USA, Federal

EARLY 2024

ARTIFICIAL INTELLIGENCE ACT

Comprehensive regulation classifying AI systems by risk levels and imposing corresponding regulatory obligations.

European Union

OCTOBER 2024

AI SEXUAL IMAGERY PROTECTION

Laws making AI-generated sexual imagery of minors a felony, even if images are not of real children.

California

JANUARY 2023

DEEP SYNTHESIS PROVISIONS

Regulations requiring clear disclosure of deepfake content use and establishing accountability for providers and users.

China, International

MARCH 2024

ELVIS ACT

Law protecting artists from unauthorized AI-generated reproductions of their voices and likenesses.

Tennessee

SEPTEMBER 2024

ELECTION DEEPFAKE LAWS

Laws prohibiting AI-generated content in political advertisements near Election Day.

California

Technical Countermeasures and Response Framework

Enterprises should adopt layered approaches to synthetic media detection. According to NIST's framework, effective detection requires analysis across multiple dimensions. Network-level monitoring must be combined with content analysis and behavioral pattern recognition. Organizations have begun implementing real-time verification protocols that examine both technical and human factors during video interactions. Subtle inconsistencies in audio-visual synchronization and behavioral patterns can reveal synthetic content.

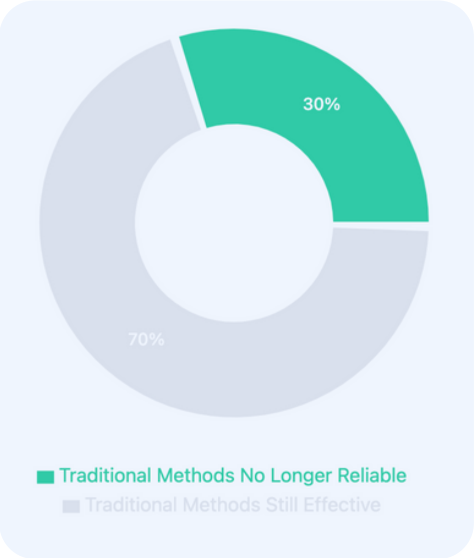
The financial sector also has several promising approaches. Leading institutions now employ advanced biometric systems that analyze multiple factors simultaneously - facial movements, voice patterns, and behavioral characteristics. These systems are particularly effective because they examine the consistency of human behavior across different communication channels, making them harder to defeat with synthetic content.

Recent incidents highlight the importance of a coordinated organizational response to synthetic media threats. Successful defenses typically involve three key components: technical detection capabilities, human awareness, and incident response protocols.

Looking Ahead

The synthetic media threat landscape presents a fundamental challenge to enterprise security. Organizations must recognize that traditional security controls no longer provide adequate protection against sophisticated deepfake attacks. The implementation of advanced AI-powered detection systems, combined with enhanced human awareness and robust verification procedures, has become critical for maintaining operational security and trust in digital interactions.

Emerging Threats



The synthetic media threat landscape continues to evolve at an unprecedented pace. The Department of Homeland Security's analysis suggests that by 2025, deepfake technology will be sufficiently advanced that 30% of enterprises will no longer consider traditional identity verification solutions reliable when used in isolation. This projection has profound implications for how organizations approach authentication and trust in digital interactions.

Recent incidents highlight the increasing sophistication of these attacks. Attackers are now combining multiple synthetic media elements - video, audio, and forged documents - in coordinated campaigns. These multi-modal attacks present particular challenges for detection systems, as they require simultaneous analysis across multiple channels while maintaining real-time performance.

Future Considerations

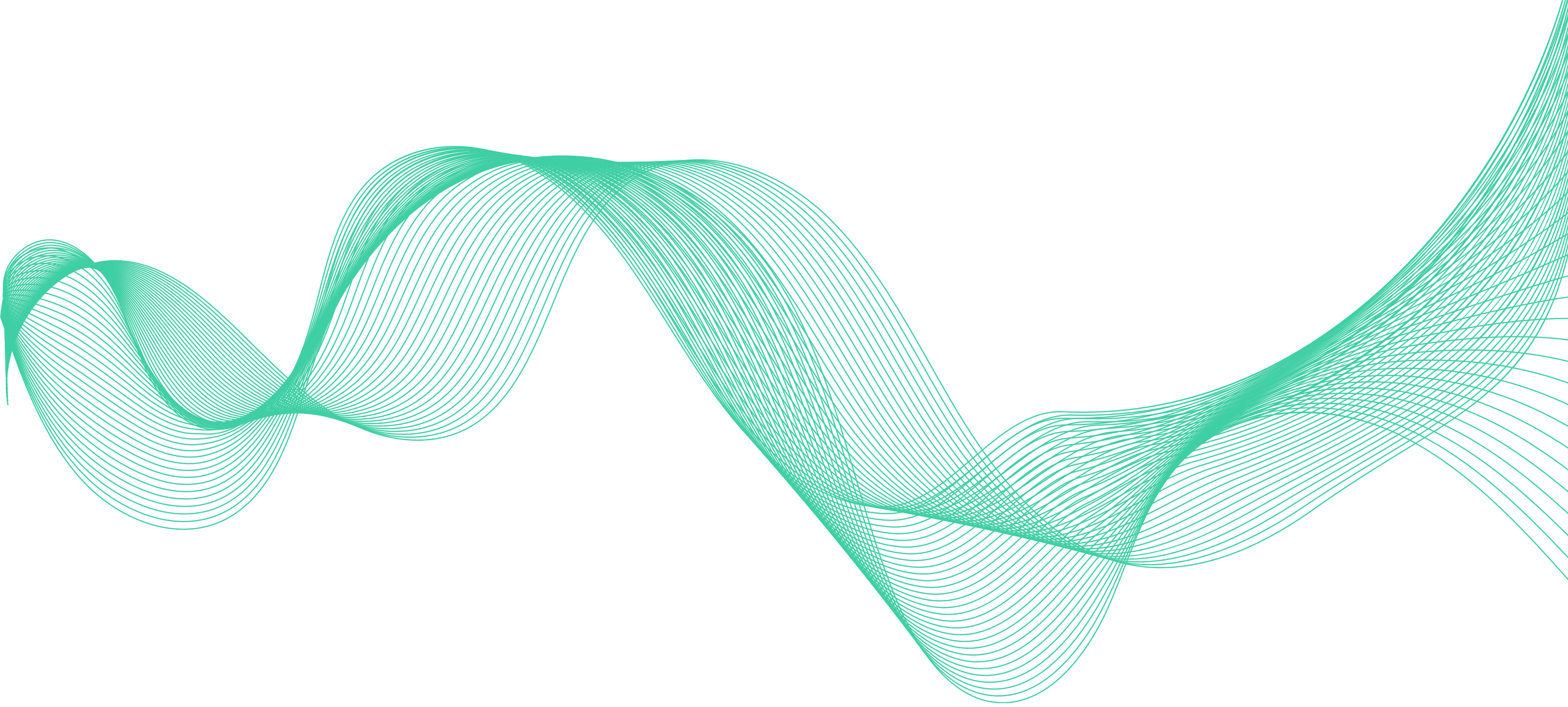
Looking ahead, organizations must prepare for increasingly sophisticated synthetic media threats. Based on current trend analysis from FS-ISAC and other security organizations, several key recommendations emerge:

First, enterprises must develop comprehensive detection strategies that combine technical controls with human oversight. This hybrid approach can effectively identify sophisticated deepfake attempts that might defeat purely technical controls.

Second, organizations should implement robust identity verification protocols that extend beyond single-point authentication. Successful attacks often exploit gaps between different verification systems. The need for integrated, cross-channel authentication approaches.

Third, incident response plans must be updated to specifically address synthetic media threats. The speed and sophistication of modern attacks require predetermined response protocols that can be activated quickly when synthetic content is detected.

Finally, organizations must invest in ongoing training and awareness programs. Well-trained employees represent a crucial line of defense against synthetic media attacks.



Conclusion

The synthetic media threat landscape presents unprecedented challenges for enterprise security. As documented across multiple incidents in 2024, organizations face increasingly sophisticated attacks that combine deepfake technology with traditional social engineering tactics. Success in countering these threats requires a comprehensive approach that combines technical controls, human awareness, and robust organizational processes.

The financial impact of failing to address these threats can be severe, as evidenced by the $25 million fraud case highlighted earlier. However, the broader implications for organizational trust and operational integrity may be even more significant. As synthetic media technology continues to evolve, organizations must remain vigilant and adaptive in their defense strategies.

About the Author

Magnus Solberg is a cybersecurity product leader specializing in AI-enabled synthetic media detection. With ~10 years of combined experience in finance and technology, he works closely with public and private sector organizations on AI safety initiatives. Previously, he covered AI and cybersecurity at J.P. Morgan and led product operations in a media-tech company focused on secure media content delivery. He holds an MBA from Chicago Booth and a Master's in Physics.

References:

1. Financial Crimes Enforcement Network. (2024). "FinCEN Alert on Fraud Schemes Involving Deepfake Media Targeting Financial Institutions." FIN-2024-Alert004.
2. FS-ISAC. (2024). "Deepfakes in the Financial Sector: Understanding the Threats, Managing the Risks." February 2024.
3. U.S. Department of State, Treasury, and FBI. (2022). "Guidance on the Democratic People's Republic of Korea Information Technology Workers." Joint Advisory, May 2022.
4. National Institute of Standards and Technology. (2024). "Reducing Risks Posed by Synthetic Content." NIST AI 100-4, November 2024.
5. California State Legislature. (2024). "California AI Transparency Act." Senate Bill No. 942, September 2024.
6. National Security Agency, FBI, & CISA. (2023). "Contextualizing Deepfake Threats to Organizations." Joint Cybersecurity Information Sheet, September 2023.
7. MITRE Corporation. (2024). "Response to NIST RFI on Sections 4.1, 4.5, and 11 of the EO Concerning Artificial Intelligence." February 2024.

RESEMBLE.AI

Contact Us

detect@resemble.ai
+1 650-822-3766
www.resemble.ai

About Resemble AI
The All-in-One AI Voice Platform.
Resemble AI delivers a cutting-edge AI Voice Generator and robust Deepfake Audio Detection, engineered for enterprises prioritizing advanced security and safety.