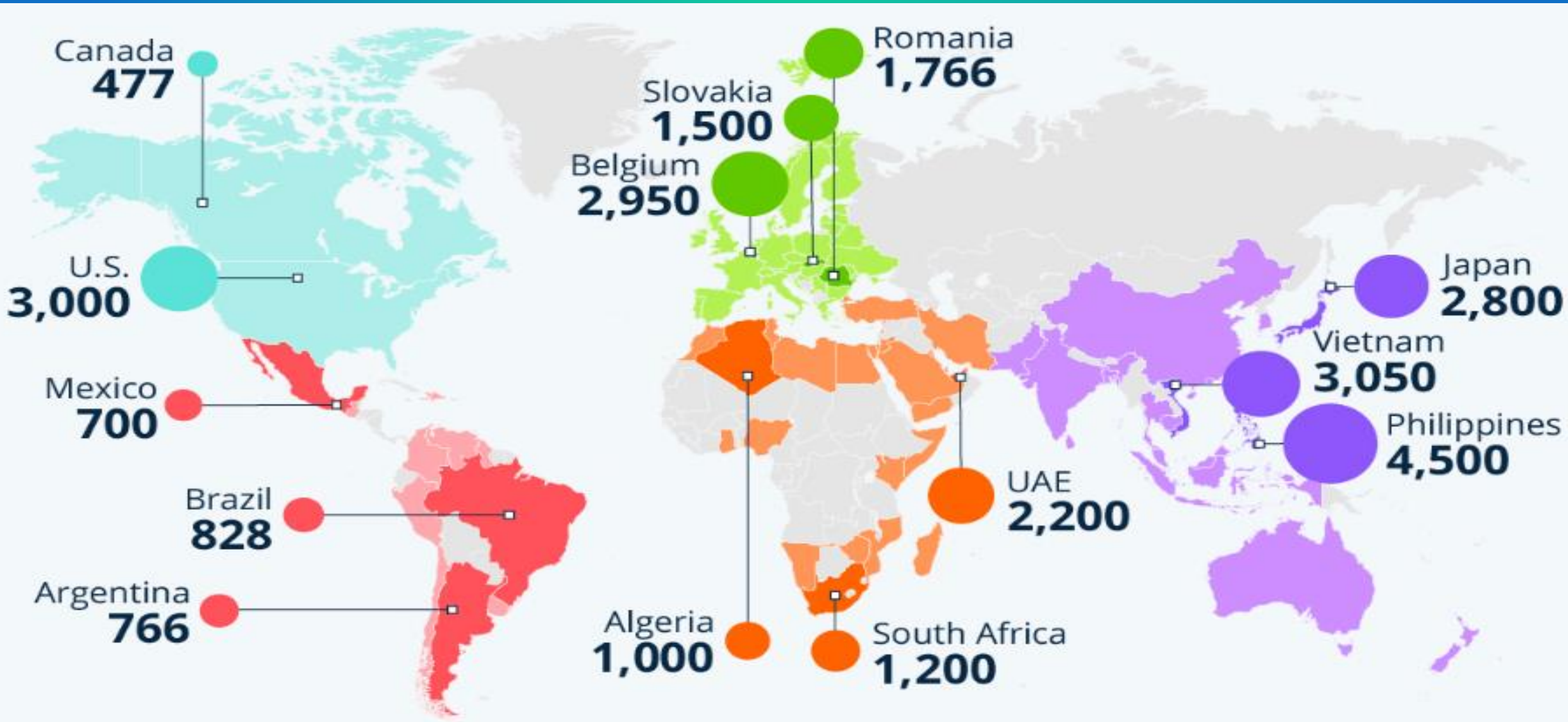
SAFEGUARDING INTEGRITY IN THE VIRTUAL WORLD – RESPONDING TO DEEP-FAKES AND INSIDER THREAT-DRIVEN FRAUD

Bernard Arinaitwe
MScIT, CFE,CISA, ISO 27001:2022 Lead Auditor, ISO 31000 Lead Risk Manager, Certified EnCE Examiner, CEH,CHFI, PECB Certified Trainer, Certified Artificial Intelligence Professional (CAIP)

The Virtual World: New Opportunities, New Threats

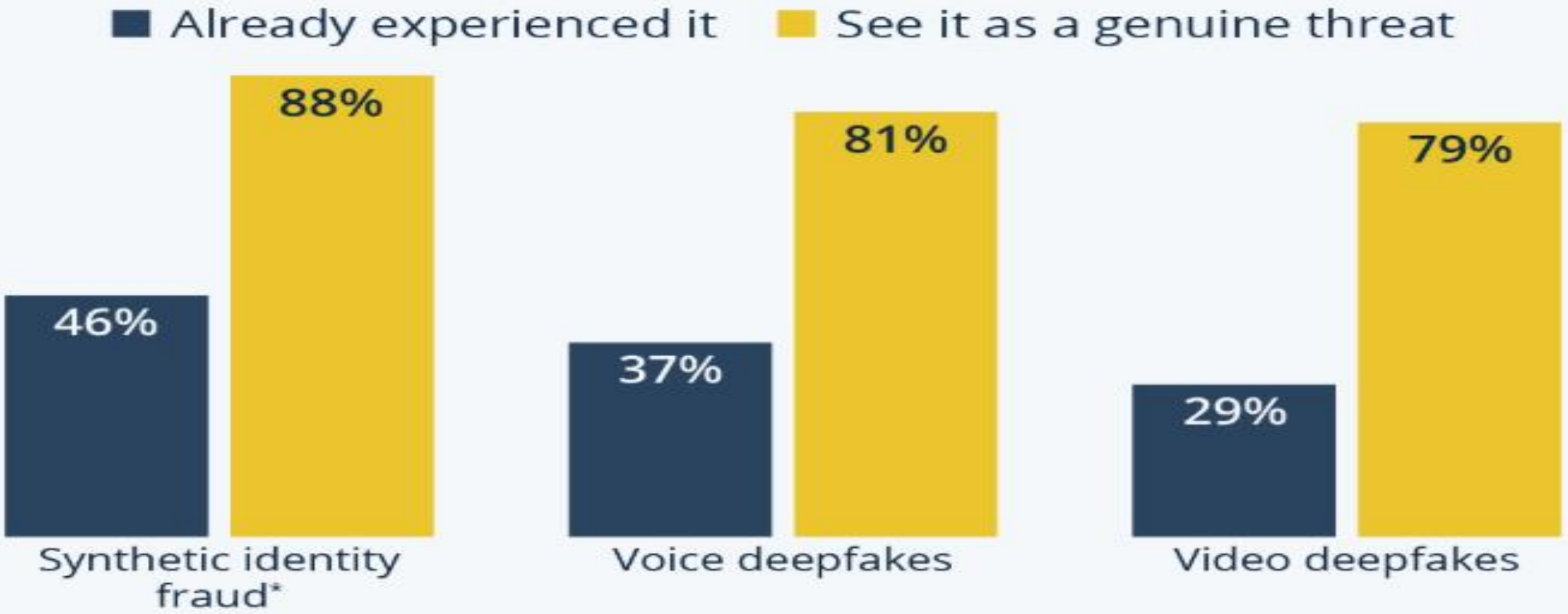
- **Digital transformation** has created boundless opportunities but also novel vulnerabilities.
- **Deep-fakes**—AI-generated synthetic media, can convincingly impersonate voices, faces, and even behaviors, making traditional verification methods obsolete.
- **Insider threats**, often overlooked, remain among the most damaging sources of fraud, especially when combined with sophisticated digital deception.

The Explosive Growth of AI-Powered Fraud



The Rise of Technology-Assisted Identity Fraud

Share of experts with the following experiences/views on advanced methods of identity fraud



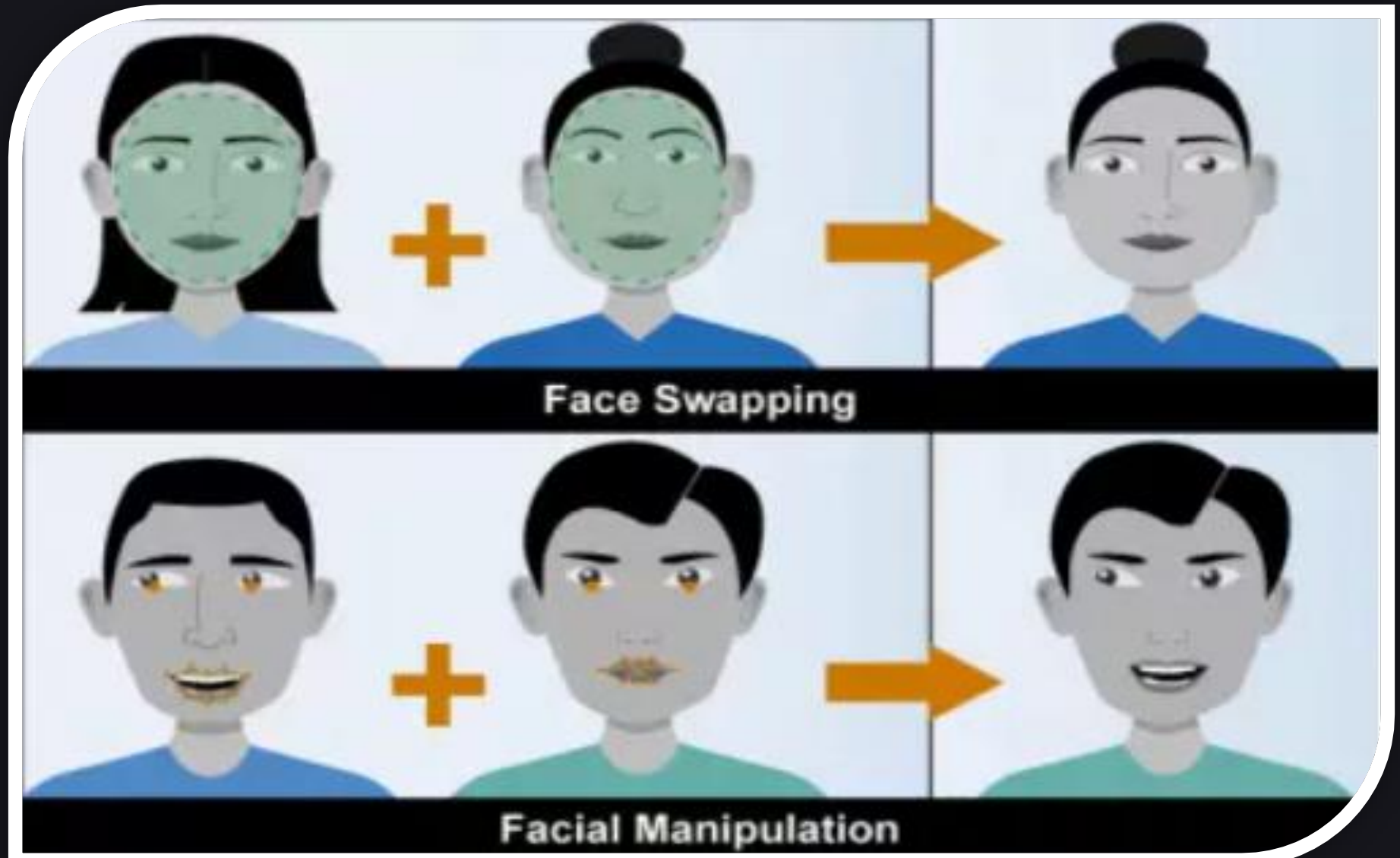
What is Deepfake?

Deepfake is a technique that uses deep learning algorithms to create fake media usually by swapping a person's face/audio from a source into another person's face in a target.

These hyper-realistic digital manipulations of images, voices, and videos have the potential to create widespread misinformation, erode trust in media, and even incite political and social unrest.

Deepfakes can be used to make people say things they never said, do things they never did, and create scenarios that never occurred.

One might recognize deepfake technology from Photoshop or mobile applications that can alter images and backgrounds, make people younger or older, mimic or clone voices, inject people into movie scenes or morph faces on the bodies of celebrities.


Face Swapping
Facial Manipulation

Deep-fakes: The New Face of Fraud

- **Implications for Uganda and Africa:**

  As digital adoption accelerates, so do targeted social engineering attacks leveraging local languages and contexts.

What are deepfakes used for?

While deepfakes offer creative and educational possibilities, they also pose serious risks when misused. Some common applications include:

Fraud and Scams: Deepfakes can be used for fraudulent activities, such as impersonating someone for financial gain or committing identity theft.

Misinformation and Fake News: One of the most concerning applications is the creation of deceptive content, such as spreading false information, defaming individuals, or manipulating public opinion.

Political Manipulation: They can be employed to create misleading political propaganda, manipulate elections, or damage the reputation of public figures.

Entertainment: Deepfakes have been used to create humorous videos or mimic famous personalities for entertainment purposes.

Creative Projects: They are sometimes utilized in filmmaking or other creative endeavors to bring fictional characters to life or enhance visual effects.

What are risks associated with deepfakes?

Deepfakes pose significant risks and ethical concerns due to their potential to deceive and manipulate.

Some reasons why they are problematic include:

Misinformation: Deepfakes can spread false information and contribute to the erosion of trust in media and information sources.

Privacy Concerns: They can infringe upon individuals' privacy rights by superimposing their likeness into compromising or fabricated situations.

Identity Theft: Deepfakes can be used to impersonate individuals, leading to identity theft, reputational damage, or extortion.

Manipulation of Reality: They blur the line between truth and fiction, making it difficult to discern genuine content from manipulated content.

Potential for Harm: Deepfakes have the potential to cause harm by inciting violence, damaging relationships, or undermining democratic processes.

BBC

Trump supporters target black voters with faked AI images

4 March 2024

Share Save

Marianna Spring
BBC Panorama and Americast

This image, created by a radio host and his team using AI, is one of dozens of fakes portraying black Trump supporters

The Daily Telegraph
NEWSPAPER OF THE YEAR

Trump to sue BBC for $1bn

President demands apology, retraction and payout by Friday if broadcaster wants to avoid legal action over biased reports

By Robert Mendick, Rob Crilly *and* Gordon Rayner

DONALD TRUMP has made an extraordinary threat to sue the BBC for $1bn, accusing it of a "reckless disregard for the truth".

The US president turned the screw on the corporation as it was reeling from the resignation of Tim Davie, its director-general, over the selective editing of Mr Trump's speech on the day of the Capitol Hill riot.

A senior official in the Trump administration suggested it "may consider" removing the BBC's White House passes as a result of the row.

The Telegraph disclosed last Monday that *Panorama* had spliced together two parts of the speech in a way that wrongly suggested Mr Trump had urged his supporters to behave violently. The revelations were contained in a leaked internal memo that also identified anti-Israel and pro-trans bias.

Lawyers acting for Mr Trump sent a legal letter to the BBC on Sunday demanding compensation, an apology and a retraction of "false" and "inflammatory statements" made about him in the *Panorama* documentary.

Mr Trump's lawyers said a "failure to comply will leave President Trump with no choice" but to pursue the broadcaster for damages for the "overwhelming financial and reputational harm". Mr Trump issued the BBC a deadline of this Friday or it would face a legal claim for $1bn (£760m) in the Florida courts.

The ultimatum was delivered as the BBC grappled with the fallout from the resignations of Mr Davie and Deborah Turness, the chief executive in charge

dent Trump by intentionally and deceitfully editing its documentary in order to try and interfere in the presidential election. President Trump will continue to hold accountable those who traffic in lies, deception and fake news."

In the letter, Mr Trump's lawyers demanded that the misleading claims be taken down, along with an apology and appropriate compensation for the "harm caused". Mr Trump's lawyers gave the corporation until 5pm on Friday to take action.

If the corporation failed to respond by then, "President Trump will be left with no alternative but to enforce his legal and equitable rights, all of which are expressly reserved and are not waived, including by filing legal action for no less than $1bn in damages". The letter concludes: "The BBC is on notice."

> @realDonaldTrump
> The TOP people in the BBC, including TIM DAVIE, the BOSS, are all quitting/FIRED, because they were caught "doctoring" my very good (PERFECT!) speech of January 6th. Thank you to The Telegraph for exposing these Corrupt "Journalists."

In June, the BBC launched a paywall for US visitors to its website, promising to deliver trustworthy news. The legal action could put these plans in jeopardy at a time when the broadcaster is seeking alternative revenue streams.

The legal threat intensifies the war of words being waged by the White House against the BBC. On Saturday, the White House condemned the corporation as a "Leftist propaganda machine".

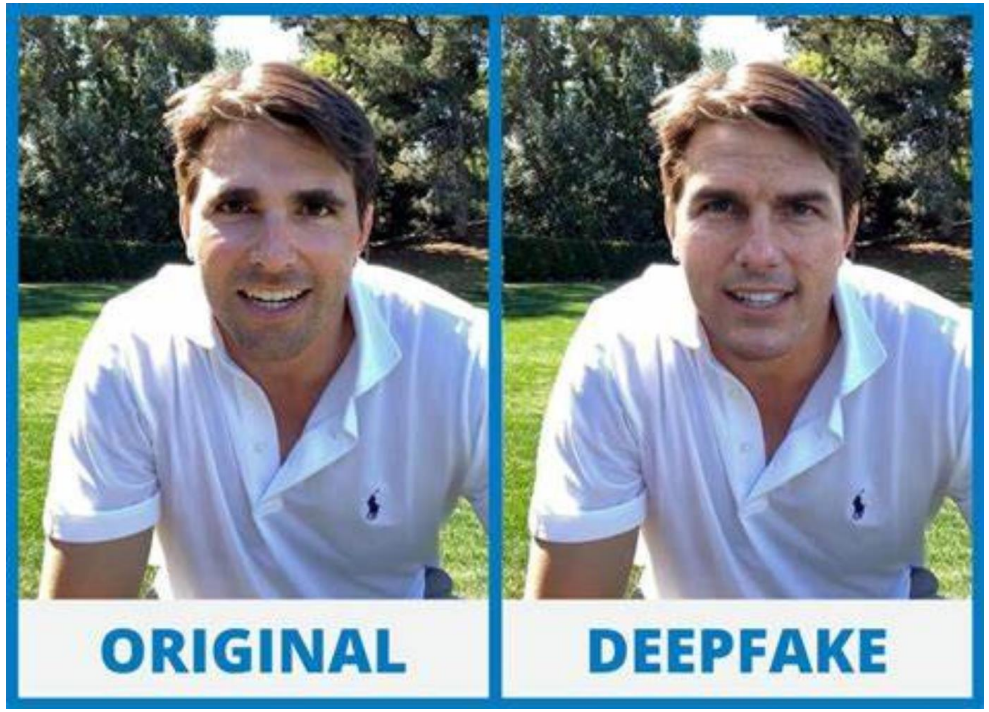
Hopes that the resignations of Mr Davie and Ms Turness would draw a line under the affair dissipated when Mr Trump took to Truth Social, his social media platform, to celebrate their resig-


TOM CRUISE
MISSION: IMPOSSIBLE
DEAD RECKONING
PART ONE
CERTIFIED
FRESH
Rotten Tomatoes
"THE BIGGEST AND BEST ACTION MOVIE OF THE YEAR"

What is Deepfake Phising?

One might recognize deepfake technology from Photoshop or mobile applications that can alter images and backgrounds, make people younger or older, mimic or clone voices, inject people into movie scenes or morph faces on the bodies of celebrities.

It exploits trust, bypasses traditional security measures, and manipulates individuals into revealing confidential information.



How does deepfake phishing work?

Deepfake phishing follows the same core principle that social engineering attacks follow, which is to confuse or manipulate users, exploit their trust and bypass traditional security measures.

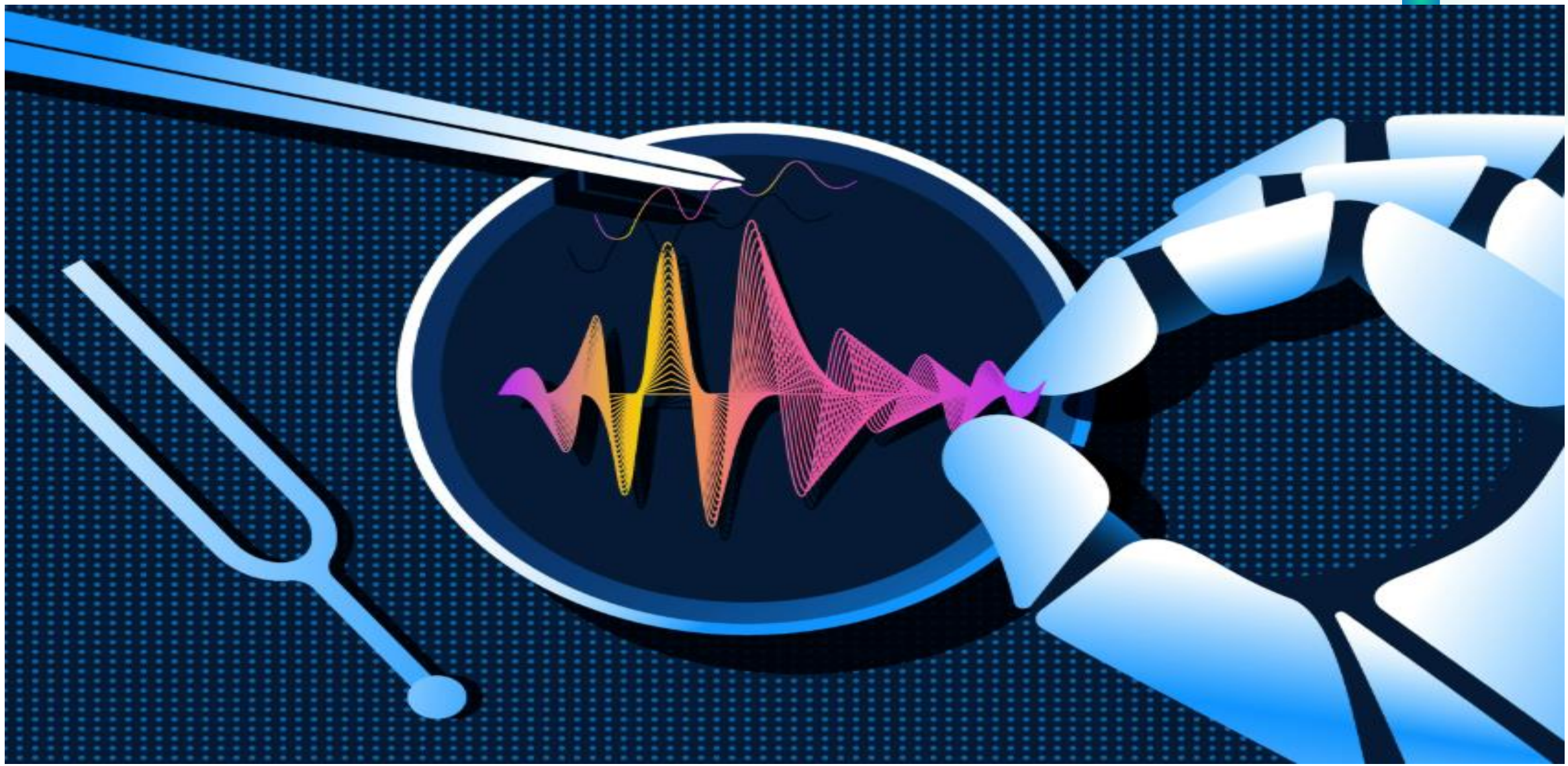
Attackers can weaponize deepfakes for phishing attacks in several different ways.

1.Emails Or Messages
Businesses are already losing billions of dollars to business email compromise (BEC) attacks (or CEO fraud) every year. Deepfakes make BEC attacks even more dangerous because threat actors can use this approach to personalize messages and make their identities seem more credible. For instance, attackers can create fake LinkedIn profiles of CEOs and use them to lure employees.

2. Video Calls
An attacker can use a video deepfake over a Zoom call to engage and convince victims to share confidential information (such as credentials) or manipulate them into carrying out unauthorized financial transactions. For example, a scammer in China used face-swapping technology to impersonate someone and convinced the victim to transfer $622,000.

3. Voice Messages
It’s extremely easy to clone anyone’s voice these days. All that’s required is a three-second clip. These deepfakes can be used to engage people in live conversations, further blurring the lines between reality and deception.

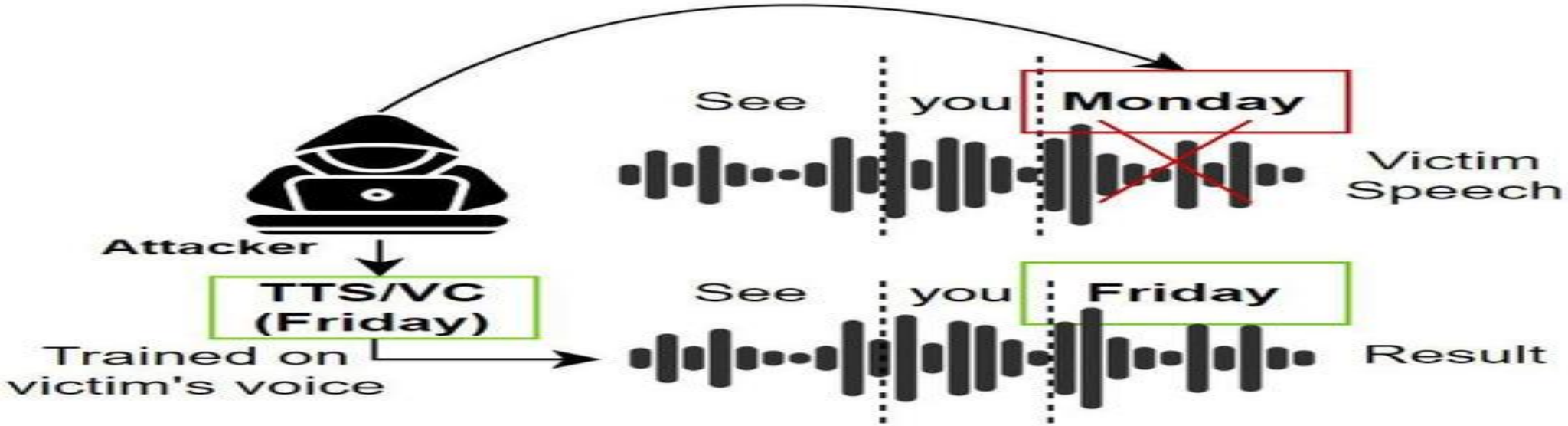
AUDIO DEEPFAKES



What is deep fake voice?

- A voice deepfake is one that closely mimics a real person's voice. The voice can accurately replicate tonality, accents, and other unique characteristics of the target person. People use AI and robust computing power to generate such voice clones or synthetic voices.

Partial Fake Speech Attacks in the Real World Using Deepfake Audio



Why should organizations be concerned about deepfake phishing?

1. It’s a fast-growing threat.

Deepfake technology is becoming increasingly sophisticated and accessible thanks to generative AI tools. Instances of deepfake phishing and fraud have surged by an astounding 3,000% in 2023.

2. It’s highly targeted.

Attackers can use deepfakes to create highly personalized attacks, targeting people based on their specific interests, hobbies and network of friends, exploiting vulnerabilities that are unique to select individuals and organizations.

3. It’s difficult to detect.

AI can mimic someone’s writing style, clone voices with near-perfect accuracy and create AI-generated faces that are indistinguishable from human faces. This makes deepfake phishing attacks extremely hard to detect.

How can organizations mitigate the risk of deepfake phishing?

- Deepfake phishing attacks are just getting started. Below are some best practices organizations can adopt to mitigate the risk of deepfake phishing:

Action Plan for Organizations

1. Improve staff awareness of synthetic content.
2. Train employees to recognize and report deepfakes.
3. Deploy robust authentication methods to reduce the risk of identity fraud.

1 Improve staff awareness of synthetic content.

Employees must be made aware of the increasing proliferation and distribution of synthetic content. Employees must be taught not to trust an online persona, individual or identity simply because they have some videos, photos or audio clips on their online profile.

2 Train employees to recognize and report deepfakes.

When it comes to phishing prevention and detection, nothing is more powerful than human intuition. Training employees to recognize and report fake online identities, visual anomalies such as lip sync inconsistencies, jerky head and torso movements, unusual audio cues, and irregular or suspicious requests is paramount. Organizations that do not have this training expertise can also consider phishing simulation programs that use real, in-the-wild example social engineering scripts.

3 Deploy robust authentication methods to reduce the risk of identity fraud.

Deploying tools such as phishing-resistant multi-factor authentication and zero-trust can certainly reduce the risk of identity theft and lateral movement to a degree. That said, security leaders can expect to see attackers attempting to hack or bypass authentication systems using clever deepfake-based social engineering techniques.

Protecting yourself from deepfake phishing

🔐 1. Be Skeptical of Urgency.

2. Verify Voice/Video Requests.

📧 3. Double-Check Email & Social Media.

🔐 4. Use 2-Factor Authentication (2FA)

🔍 5. Look for Deepfake Signs

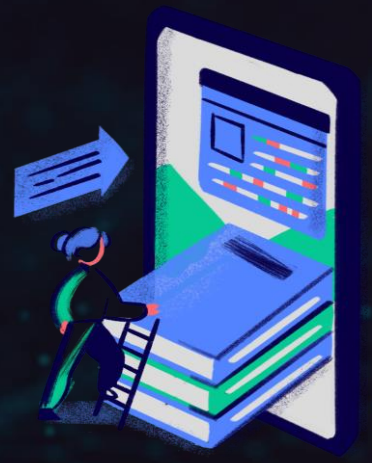
6. Use Verification Tools.

7. Stay Educated.

💼 8. Train Your Team (If at Work)

Building a Cybersecurity Culture

Promoting awareness and education



Leadership’s role in emphasizing security

Establishing clear policies and procedures

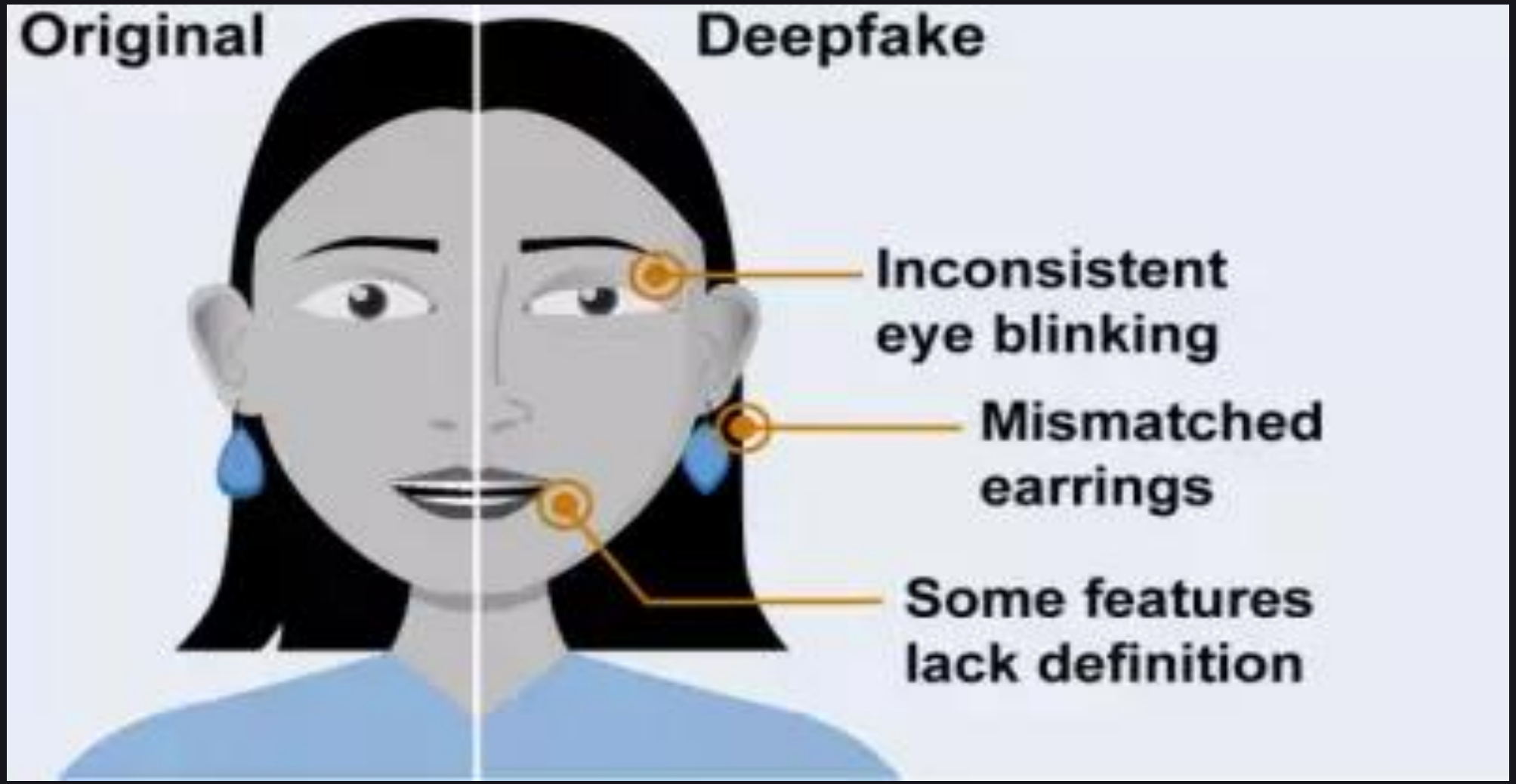
How can you spot a deepfake?

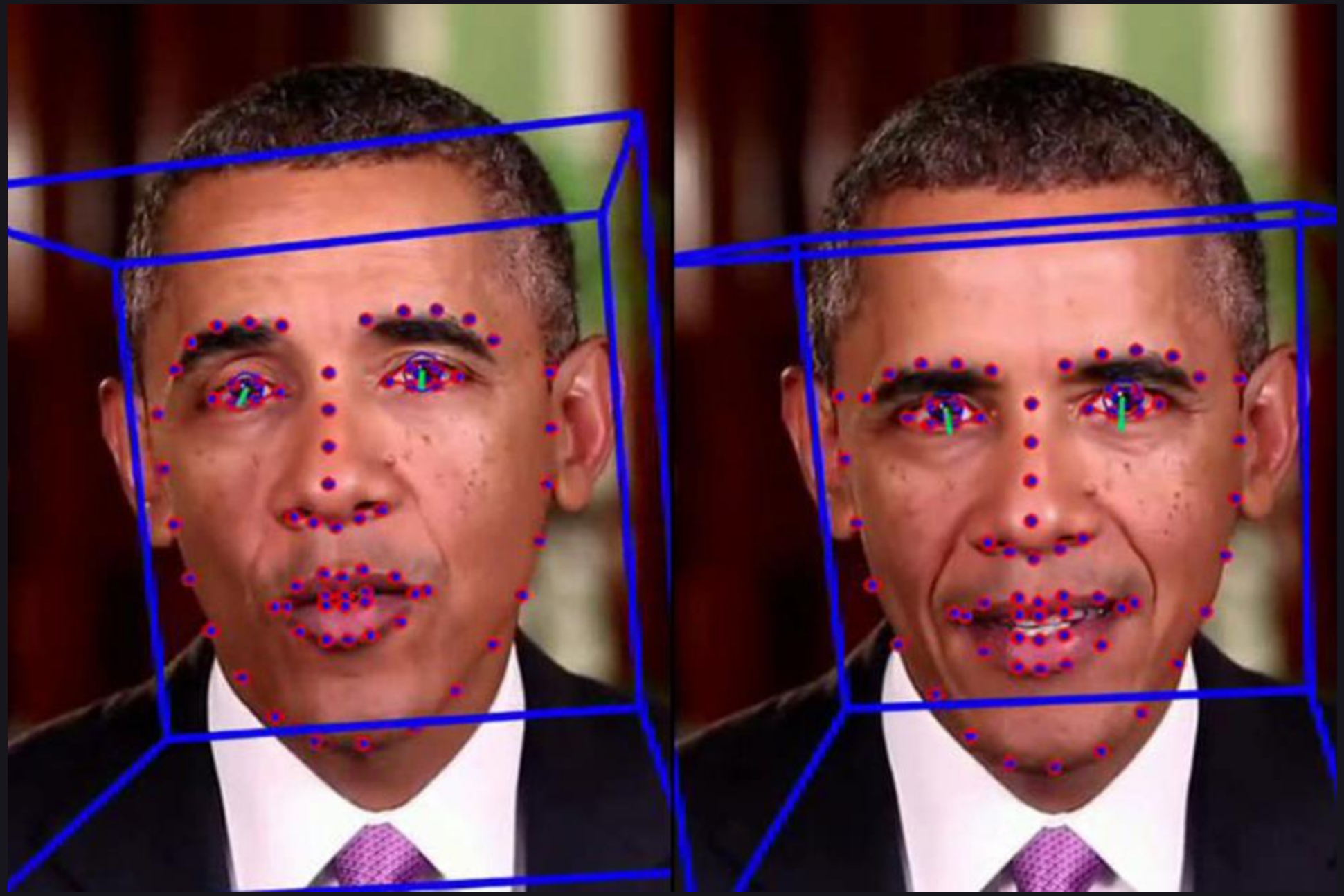
a) **Inconsistent Audio-Visual Cues**: The audio may not sync perfectly with the lip movements, or there may be discrepancies between the subject's voice and their actions.

b) **Contextual Clues**: Consider the context of the content. If it seems highly unlikely or out of character for the individual depicted, it could be a deepfake.

c) **Source Verification**: Verify the source of the content and look for signs of manipulation or tampering.

How can you spot a deepfake?





How can you spot a deepfake?

Unnatural Facial Expressions or Movements: Deepfakes may exhibit subtle glitches or unnatural movements in the subject's face or body.

Pay attention to the face. High-end deepfake manipulations are almost always facial transformations.

Pay attention to the cheeks and forehead. Does the skin appear too smooth or too wrinkly? Is the agedness of the skin similar to the agedness of the hair and eyes? Deepfakes may be incongruent on some dimensions.

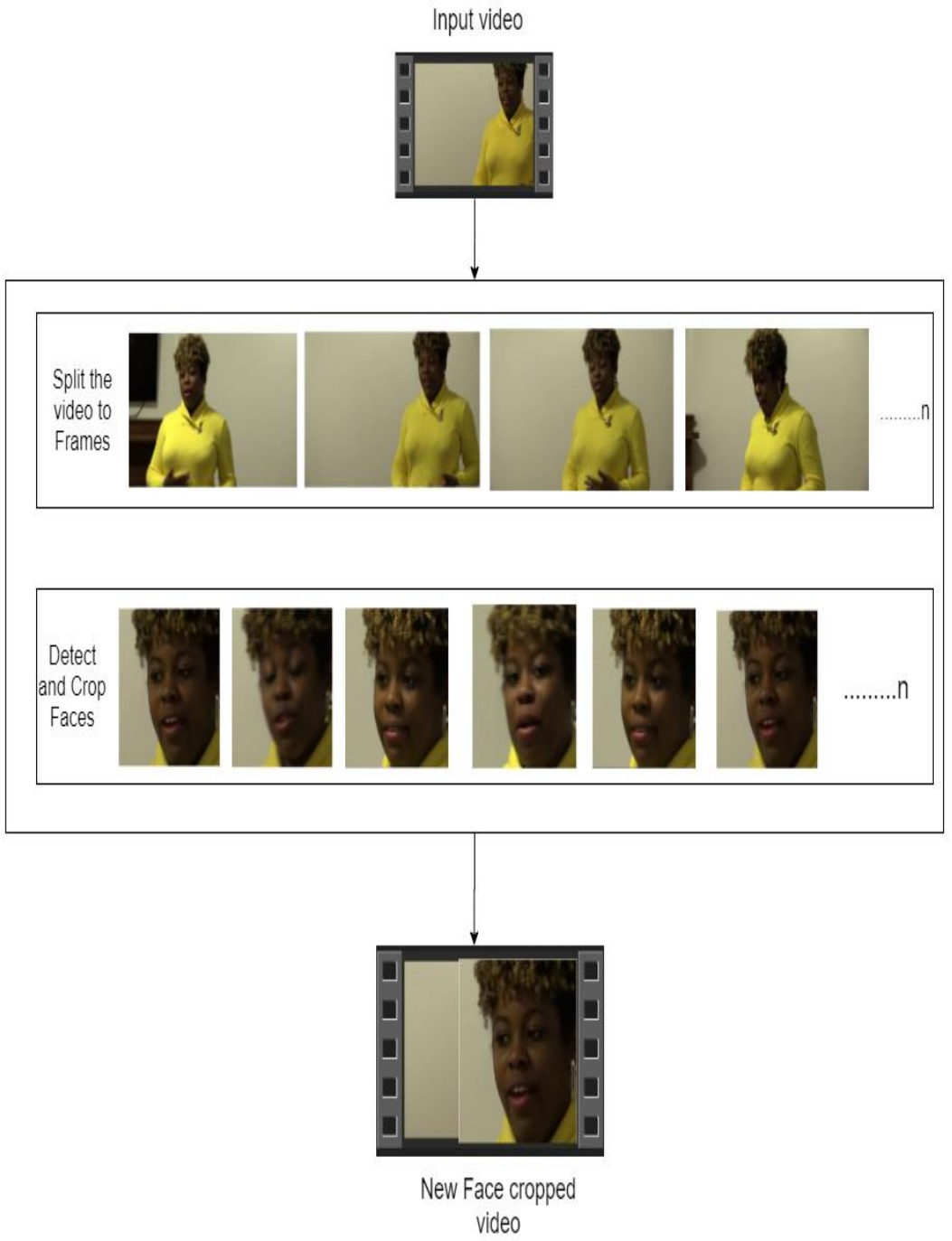
Pay attention to the eyes and eyebrows. Do shadows appear in places that you would expect? Deepfakes may fail to fully represent the natural physics of a scene.

Pay attention to the glasses. Is there any glare? Is there too much glare? Does the angle of the glare change when the person moves? Once again, deepfakes may fail to fully represent the natural physics of lighting.

Pay attention to the facial hair or lack thereof. Does this facial hair look real? Deepfakes might add or remove a mustache, sideburns, or beard; but, deepfakes may fail to make facial hair transformations fully natural.

Pre-processing

1. SPLIT VIDEO INTO FRAMES
2. FACE DETECTION
3. CROPING FACE
4. CREATING NEW FACE CROPPED VIDEO
5. SAVING THE FACE CROPPED VIDEO



How can you spot a deepfake?

Check the Source: Verify the origin of the video. If it's from a reputable source, it's more likely to be authentic.

Look for Inconsistencies: Watch for inconsistencies in lighting, shadows, and reflections that may indicate manipulation.

Audio Analysis: Pay attention to audio quality and consistency. Mismatched or unnatural sounds could be a sign of manipulation.

Frame Analysis: Analyze individual frames for anomalies, such as strange artifacts or inconsistencies that might occur during video editing.

Deepfake Detection Tools: Use specialized tools designed to detect deepfakes. These tools may use AI algorithms to identify unnatural facial expressions or inconsistencies. Reverse Image/Video Search: Use reverse image or video search engines to check if the content has been used elsewhere, indicating potential manipulation.

In spite of some legitimate applications like films and commerce, deepfakes are more commonly used for exploitation. Some studies have shown that much of deepfake content online is pornographic, and deepfake pornography disproportionately victimizes women.

There is also concern about potential growth in the use of deepfakes for disinformation. Deepfakes could be used to influence elections or incite civil unrest, or as a weapon of psychological warfare. They could also lead to disregard of legitimate evidence of wrongdoing and, more generally, undermine public trust.

What can be done to protect people?

As discussed above, researchers and internet companies, such as Microsoft and Intel, have experimented with several methods to detect deepfakes. These methods typically use AI to analyze videos for digital artifacts or details that deepfakes fail to imitate realistically, such as blinking or facial tics. But even with these interventions by tech companies, there are a number of policy questions about deepfakes that still need to be answered.

For example:

- **What can be done to educate the public about deepfakes to protect them and help them identify real from fake?**
- **What rights do individuals have to privacy when it comes to the use of deepfake technology?**

Deepfakes are powerful tools that can be used for exploitation and disinformation. With advances making them more difficult to detect, these technologies require a deeper look.

INSIDER THREATS

Insider Threats:
The Danger Inside Organizations

Uganda police crime report 2024

S/No	Offences	No. of Cases Reported					
		2024	2023	2022	2021	2020	2019
1.	Embezzlement	92	101	128	169	143	194
2.	Causing Financial loss	76	49	55	50	32	62
3.	Abuse of office	101	42	61	66	77	109
4.	Counterfeiting	209	262	388	271	238	394
5.	Forgeries & Uttering of Documents	894	868	804	747	629	911
6.	Issuing False Cheques	94	135	185	186	190	320
7.	Bank and Other Corporate Frauds	90	43	82	102	63	33
8.	Obtaining By False Pretences	10,449	10,709	10,652	8,634	8,069	10,598
9.	Cyber (Computer) crimes	474	245	286	258	253	248
10.	Prevention of Corruption Act	25	32	37	57	32	32

FRAUD SCHEMES AND DETECTION MECHANISMS

SCHEMES

ASSET MISAPPROPRIATION SCHEMES
are the most common but least costly

89% of cases

$120,000 median loss

FINANCIAL STATEMENT FRAUDS
are the least common but most costly

5% of cases

$766,000 median loss

CORRUPTION
Almost half of all reported cases included corruption

HIGHEST RISK ASSET MISAPPROPRIATION SCHEMES

DETECTION

43% of frauds were detected by tips,
which is more than 3x as many cases as the next most common method

More than HALF of tips come from employees

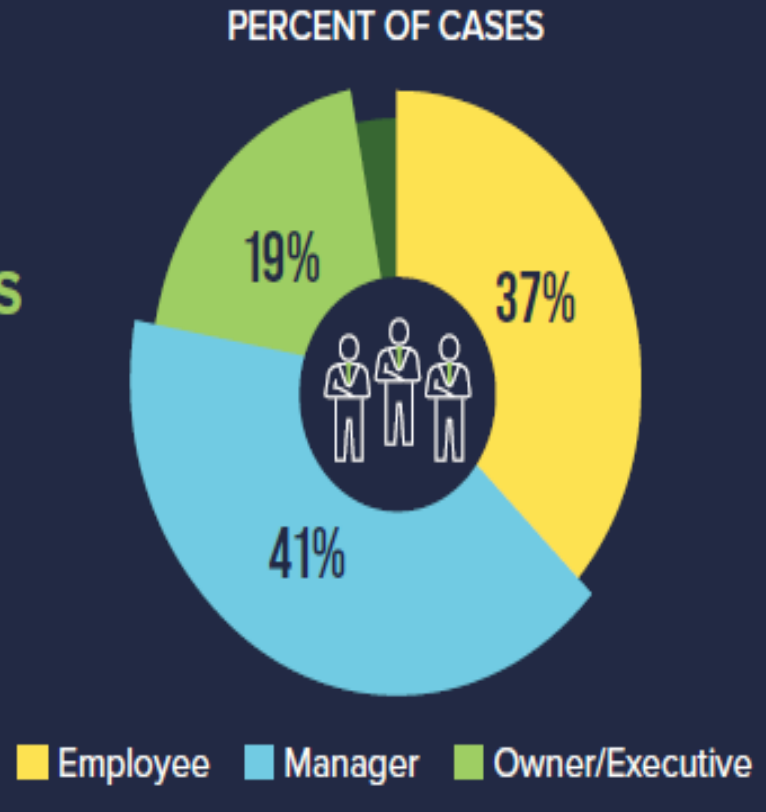
and nearly ONE-THIRD come from vendors and customers

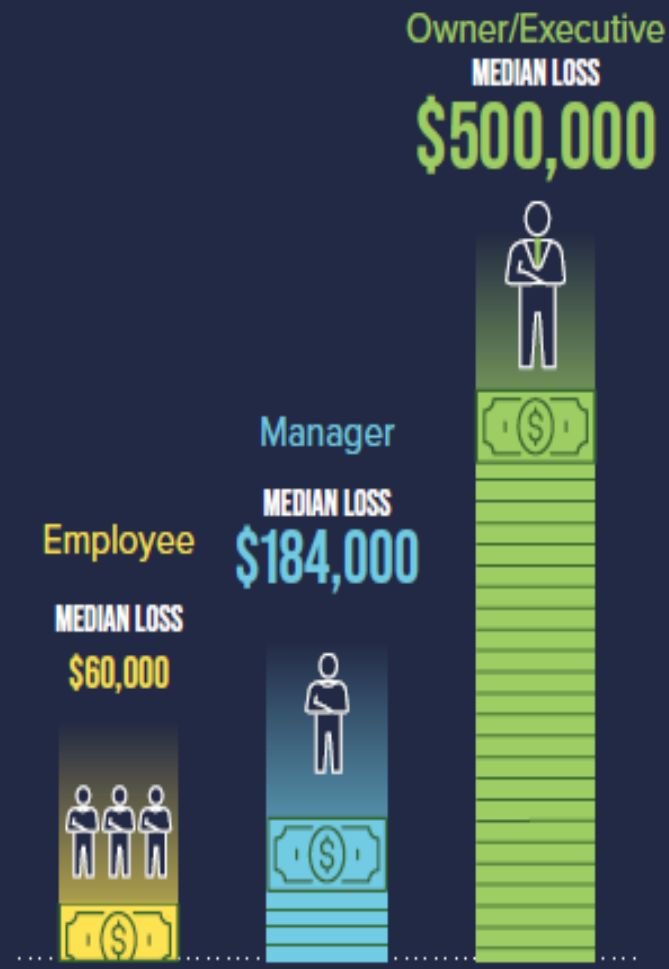
- Employee
- Customer
- Vendor

PROFILE OF A FRAUDSTER

LEVEL OF AUTHORITY

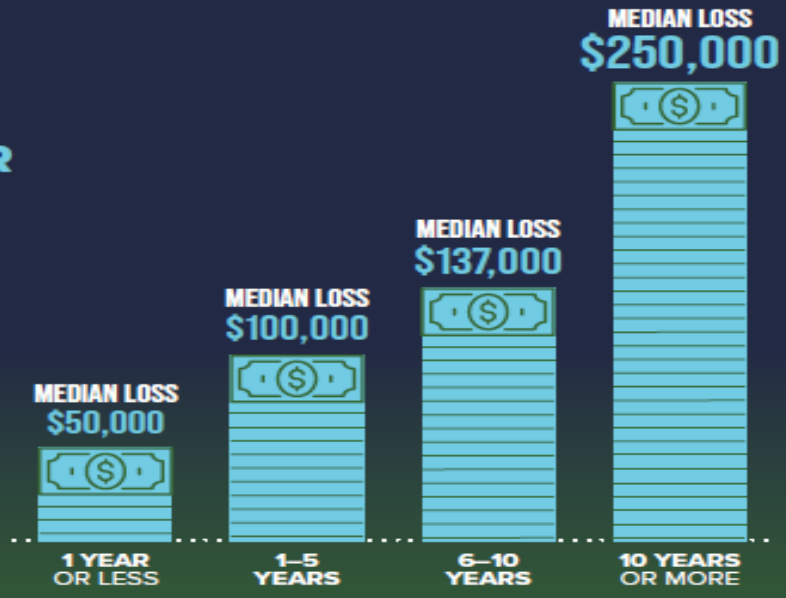
Most fraudsters were employees or managers, but **FRAUDS PERPETRATED BY OWNERS AND EXECUTIVES WERE THE COSTLIEST.**





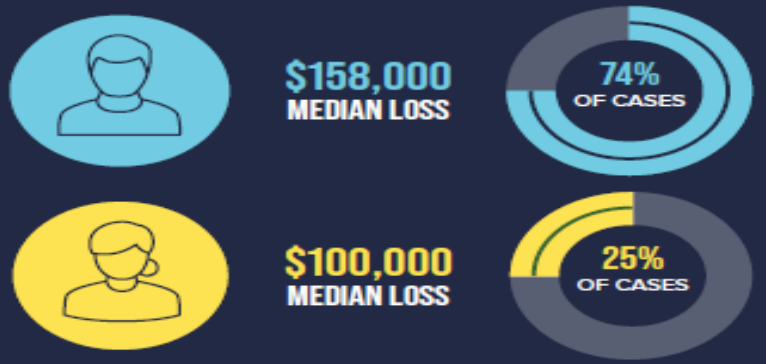
TENURE

THE LONGER a fraudster has worked for an organization, **THE MORE COSTLY** their fraud.



GENDER

WOMEN committed fewer frauds and caused lower losses.



EDUCATION

TWO-THIRDS of occupational fraudsters **HAD A UNIVERSITY DEGREE OR HIGHER.**

Fraudsters **WITHOUT A DEGREE** caused **LOWER LOSSES.**

No university degree
$100,000 MEDIAN LOSS

University degree or higher
$200,000 MEDIAN LOSS

COLLUSION

FRAUDSTERS WHO COLLUDED with others caused median losses **MORE THAN 3X AS HIGH** as those who acted alone.

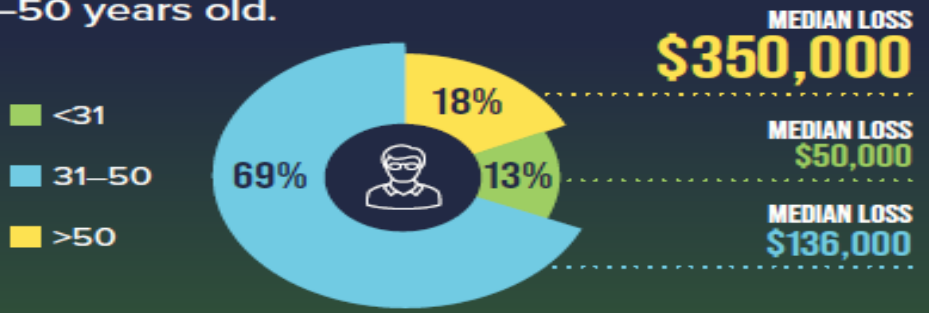
1 Perpetrator
MEDIAN LOSS
$75,000

2+ Perpetrators
MEDIAN LOSS
$250,000

AGE

MORE THAN TWO-THIRDS of fraudsters were 31–50 years old.

FRAUDSTERS OVER THE AGE OF 50 caused the highest median losses.



INSIDER THREATS IN AN ORGANIZATION

MALICIOUS INSIDERS

Employees or partners who misuse their legitimate access to confidential data for personal gain

INSIDE AGENTS

Malicious insiders recruited by external parties to steal, alter, misuse or delete confidential data

EMOTIONAL EMPLOYEES

Emotional attackers who seek to cause harm to their organization as revenge for something perceived wrongly

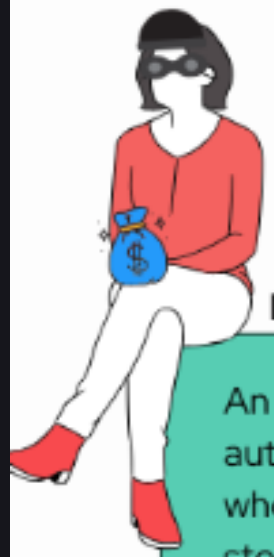
RECKLESS EMPLOYEES

Employees or partners who neglect the rules of an organization's cybersecurity policy

THIRTY-PARTY USERS

Third-party vendors who take advantage of their access to compromise the security of sensitive information

TYPES OF INSIDER THREATS



Malicious Insider

An individual with authorized access who intentionally steals, alters, or destroys data and systems for personal gain, revenge, or other harmful purposes. These individuals often have in-depth knowledge of security protocols, making their actions particularly damaging and difficult to detect.

Negligent Insider

An employee or authorized user who unintentionally compromises security through errors, carelessness, or a failure to adhere to established security policies and procedures. Examples include falling victim to phishing attacks, using weak passwords, or mishandling sensitive information.

Compromised Insider

A user whose account or credentials have been compromised by an external attacker.

The attacker then uses the insider's access to carry out malicious activities, such as data theft, malware installation, or system disruption, often without the user's knowledge.

Insider Threats To Watch Out for With a Remote Workforce

★★★★★

01
Shifts in Behavior

02
Unexplained Access Patterns

03
Unusual Data Movement

04
Policy Breaches

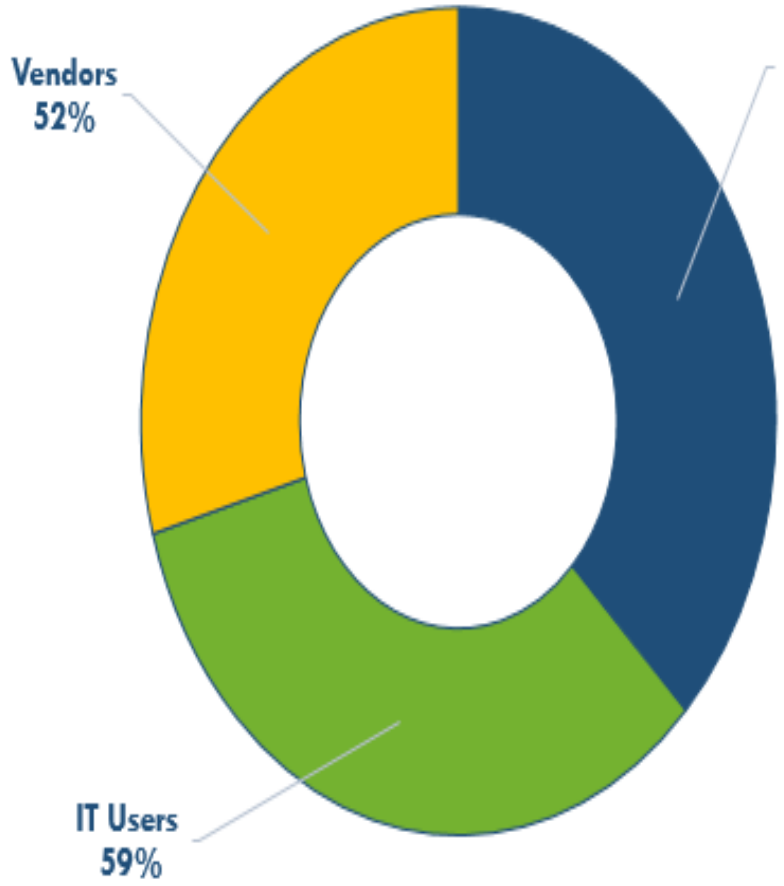
05
Privileged Access Abuse

06
Unhappy or Disengaged Employees

07
Third-Party Access Risks


scam
centers

BIGGEST INSIDER THREAT ACTORS



CONSEQUENCES OF INSIDER THREATS

1. Financial Loss

Insider threats cause massive financial damage, whether intentional or accidental. Organizations lose an average of $11.5 million annually to insider threats, with each malicious attack costing around $701,500.

Beyond direct losses, companies face expenses for incident response, forensic investigations, security upgrades, and regulatory fines under laws like GDPR and SOX. For publicly traded companies, insider threats can also shake investor confidence and drive down stock prices.

2. **Reputational Damage**

Reputational damage from insider threats is rising fast, with attacks aimed at harming a company's image jumping from 8% to 37%. Leaks, data breaches, and internal misconduct can quickly erode public trust and brand credibility.

Customer trust is often the first to go. Data breaches push consumers toward safer alternatives, and once lost, trust is costly and difficult to regain.

"It takes 20 years to build a reputation and a few minutes of cyber-incident to ruin it." Stephane Nappo, Cybersecurity Expert

3. Legal and Compliance Issues

Insider threats can trigger serious legal and compliance issuesFor industries like finance and healthcare, insider threats can result in revoked licenses and restricted operations. Enforcing strong access controls, monitoring systems, and legal safeguards is crucial for staying compliant and avoiding costly setbacks.


Insider Threats
HOW TO IDENTIFY AND PREVENT

What makes Insider threat **difficult to detect**

- Insiders have legitimate access to data
- Remote working and using personal devices
- Use of third-party applications

INSIDER ATTACK MOTIVATIONS

Organizational Politics

Pressure

Anger

Greed

Financial Gain

Treachery

Discontentment

Jealousy

Understanding User Behavior Based Insider Detection

Definition:
User Behavior Based Insider Detection leverages analytics to monitor and assess users' digital activities, identifying anomalies that may indicate insider threats or fraud.

Why Focus on User Behavior?

Insiders often have legitimate access, making traditional perimeter defenses ineffective.
Behavioral patterns can reveal subtle, early indicators of malicious intent.

Key Concepts:

Baseline Profiling: Establishing normal user activity.
Anomaly Detection: Spotting deviations from the baseline.

Techniques & Technologies

Behavioral Analytics:

- Machine learning models analyze login times, file access, data transfers, and application usage.
- Examples: Sudden large data downloads, off-hours access, unusual privilege escalations.

User and Entity Behavior Analytics (UEBA):

- Integrates data from multiple sources (logs, emails, file activity).
- Correlates behaviors across users and devices for context-aware alerts.

Continuous Monitoring:

- Real-time detection enables rapid response to suspicious activities.
- Automated risk scoring for users based on observed actions.

Challenges & Best Practices

Challenges:

- **False Positives:** Legitimate activity may appear anomalous.
- **Privacy Concerns:** Balancing detection with user privacy rights.
- **Adaptive Threats:** Malicious insiders may mimic normal behavior.

Best Practices:

- Regularly update behavioral baselines to reflect role changes.
- Combine technical controls with organizational policies (e.g., regular training, access reviews).
- Foster a culture of security awareness to encourage reporting of suspicious behavior.

CFE's Role:

- Interpret behavior analytics in context.
- Coordinate with IT and HR for thorough investigations.

WHY DO CARS HAVE BREAKS?

Conclusion

Safeguarding integrity in the virtual world is a shared responsibility.
As fraud examiners and anti-fraud professionals, our vigilance, adaptability, and commitment to continuous learning will determine our collective success. Let’s be proactive, not just reactive, in defending against both high-tech and human-driven threats.

Thank you. I look forward to your questions and a lively discussion!

Thank You

Safeguarding Integrity in the Virtual World – Responding to Deep-fakes and Insider Threat-Driven Fraud

Bernard Arinaitwe
MScIT, CFE,CISA, ISO 27001:2022 Lead Auditor, ISO 31000 Lead Risk Manager, Certified EnCE Examiner, CEH,CHFI, PECB Certified Trainer, Certified Artificial Intelligence Professional (CAIP)