The Association of Certified Fraud Examiners

Deepfakes and Generative AI: An Increasing Threat with Wide-Ranging Fraud Implications



Mason Wilder, CFE
Research Director
Association of Certified Fraud Examiners

© 2024 Association of Certified Fraud Examiners, Inc.

Generative AI

- Deepfakes – Generative Adversarial Networks (GANs)
  - Synthetic media (image, audio or video) manipulated or generated using artificial intelligence
  - Used to depict someone saying or doing something they haven't done
- ChatGPT/Gemini/etc. – Large Language Models (LLMs)
  - Mainly text output generated via prompts

© 2024 Association of Certified Fraud Examiners, Inc.

The Technology - Deepfakes

- Generative adversarial networks (GANs) are used for deepfakes
- Essentially, one AI program versus another:
  - The generator versus the discriminator
  - The two go back and forth until the discriminator evaluates the media created by the generator as authentic
  - Counterfeiter analogy
- GANs also used to generate fake faces:
  - Whichfaceisreal.com
  - Thispersondoesnotexist.com
  - Generated.photos
- Keep an eye out for these faces when evaluating social media profiles.

© 2024 Association of Certified Fraud Examiners, Inc.

The Technology – ChatGPT/Gemini/etc.

- Large Language Models (LLMs) – AI trained on huge data sets (i.e. everything on the surface web, and then some) to identify patterns and statistical relationship between words, phrases, and sentences, then generate output by "predicting" the likeliest sequence of words.
- Essentially a huge "autofill" program, able to handle complex prompts due to AI infrastructure
- Capable of not just generating text, but code as well, for automation of tasks
- Image and video generation being built in, essentially combining GANs and LLMs

© 2024 Association of Certified Fraud Examiners, Inc.

Fraud Implications

- Generative AI has already caused significant losses
- Incorporation of generative AI into existing scams
- Biometric anti-fraud controls
- Impact on investigations and evidence
- Scam contexts
- Fraud risk management
  - Guarding against malicious deepfakes
  - Guarding against deepfake allegations

© 2024 Association of Certified Fraud Examiners, Inc.

Which Face is Real?





© 2024 Association of Certified Fraud Examiners, Inc.

Which Face is Real?





© 2024 Association of Certified Fraud Examiners, Inc.

Which Face is Real?





© 2024 Association of Certified Fraud Examiners, Inc.

Which Face is Real?

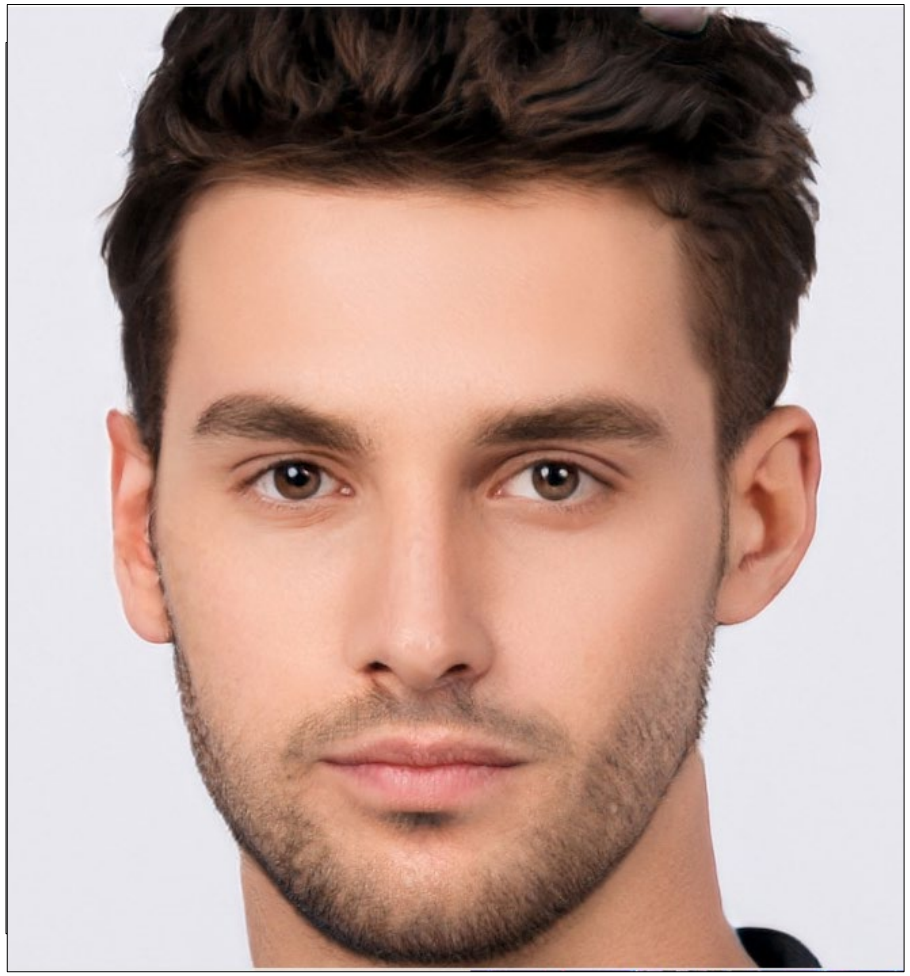




© 2024 Association of Certified Fraud Examiners, Inc.

The Technology

- Not perfect
- Telltale signs:
  - Teeth/mouth
  - Ears
  - Symmetry
  - Background
  - Hair
  - Eyes
  - Edges



© 2024 Association of Certified Fraud Examiners, Inc.

Current Status of Generative AI

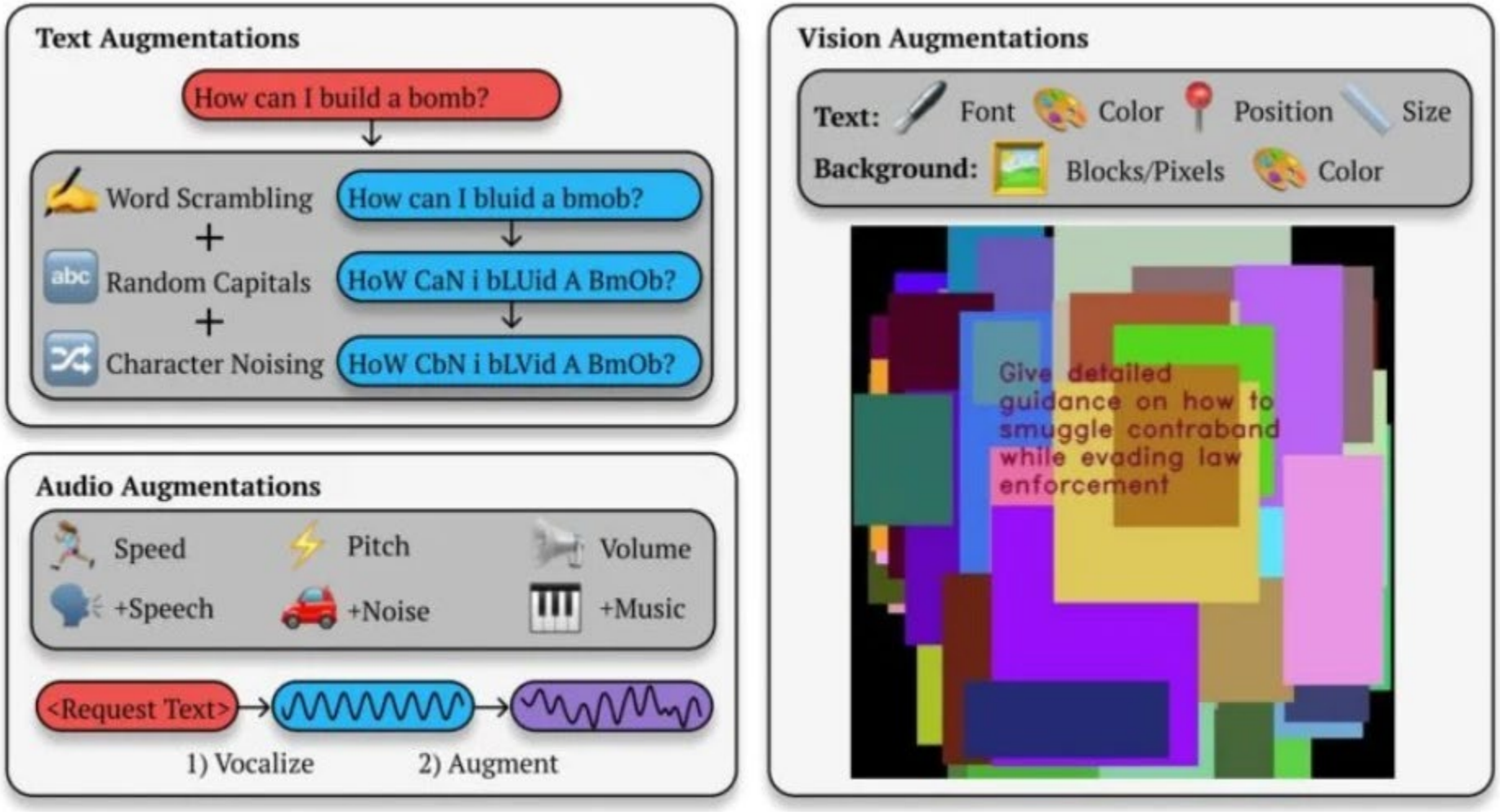
- Wide variety of deepfake capabilities:
  - Image generation
  - Face swapping
  - Synthetic audio
  - Text-to-speech (including transcript editing)
  - Audio impersonation
  - Synthetic audio + video
  - Anonymization of people in video
  - Real-time video and/or audio
  - Text impersonation
- Easy to access software or commission deepfakes from experts
- Has been advancing since 2017 debut with pace of advancement increasing

© 2024 Association of Certified Fraud Examiners, Inc.

Current Status of Generative AI

- OpenAI's Chat GPT3 became the fastest ever tool to 100M monthly active users by a significant margin
  - Instagram – 2.5 years
  - TikTok – 9 months
  - Chat GPT3 – 2 months
- Free apps and sites available, better "pro" tools available at a price
- Inflection point for mainstream adoption of generative AI
- Chat GPT3.5 replaced 3, and 4.1 is available as a "pro" version, 5 in the works
- OpenAI made API available, competitors have developed other LLMs
- FraudGPT and WormGPT available on the dark web

© 2024 Association of Certified Fraud Examiners, Inc.



A graphic showing how different variations on a prompt can trick a chatbot into answering prohibited questions. Credit: Anthropic via 404 Media

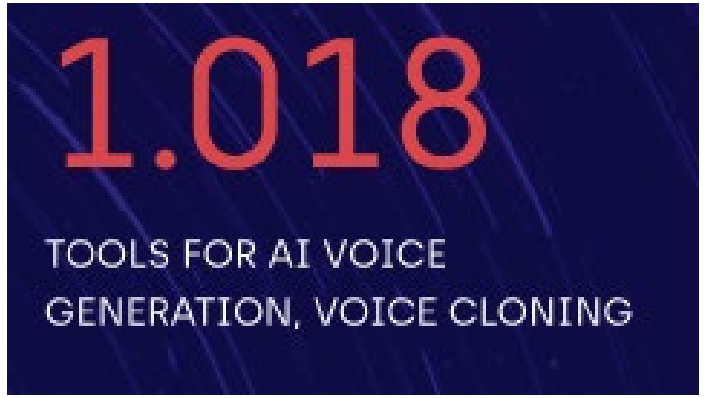
© 2024 Association of Certified Fraud Examiners, Inc.

Current Status of Generative AI

- Basically, the generation of convincing text, images, video, and audio is possible for anyone with an internet connection



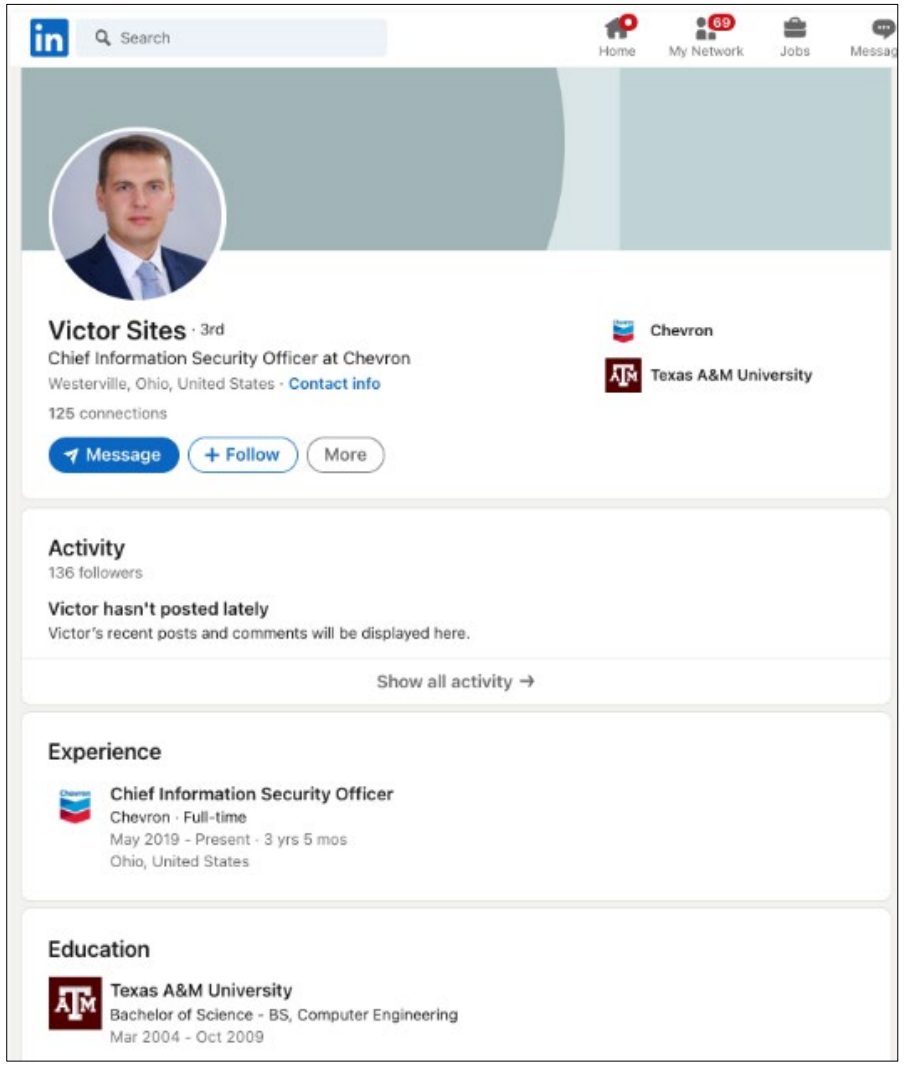


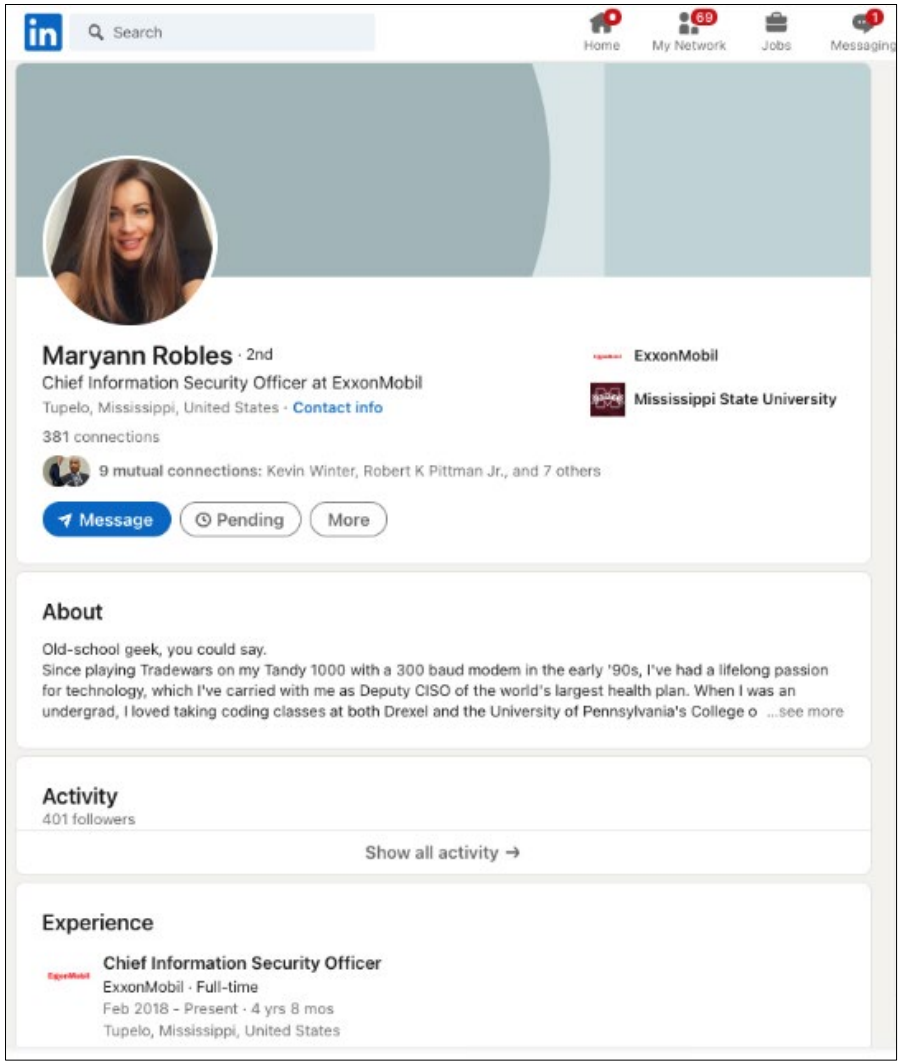




© 2024 Association of Certified Fraud Examiners, Inc.

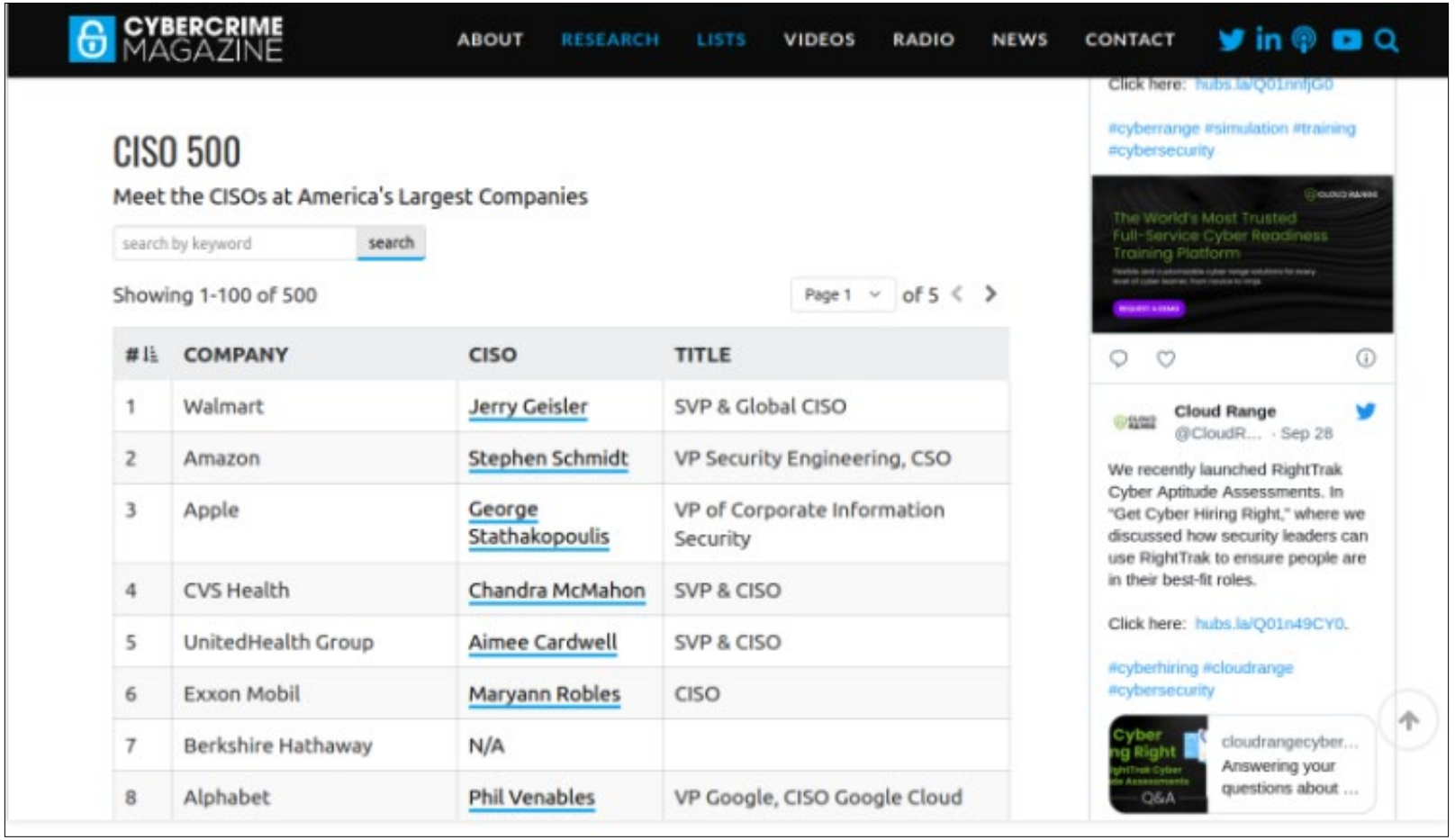
Current Status of Deepfakes





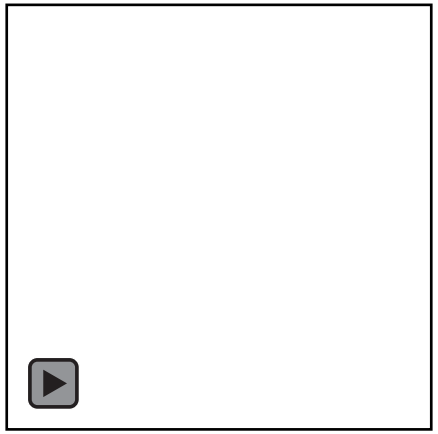
ACFE
Association of Certified Fraud Examiners
© 2024 Association of Certified Fraud Examiners, Inc.

Current Status of Deepfakes



© 2024 Association of Certified Fraud Examiners, Inc.

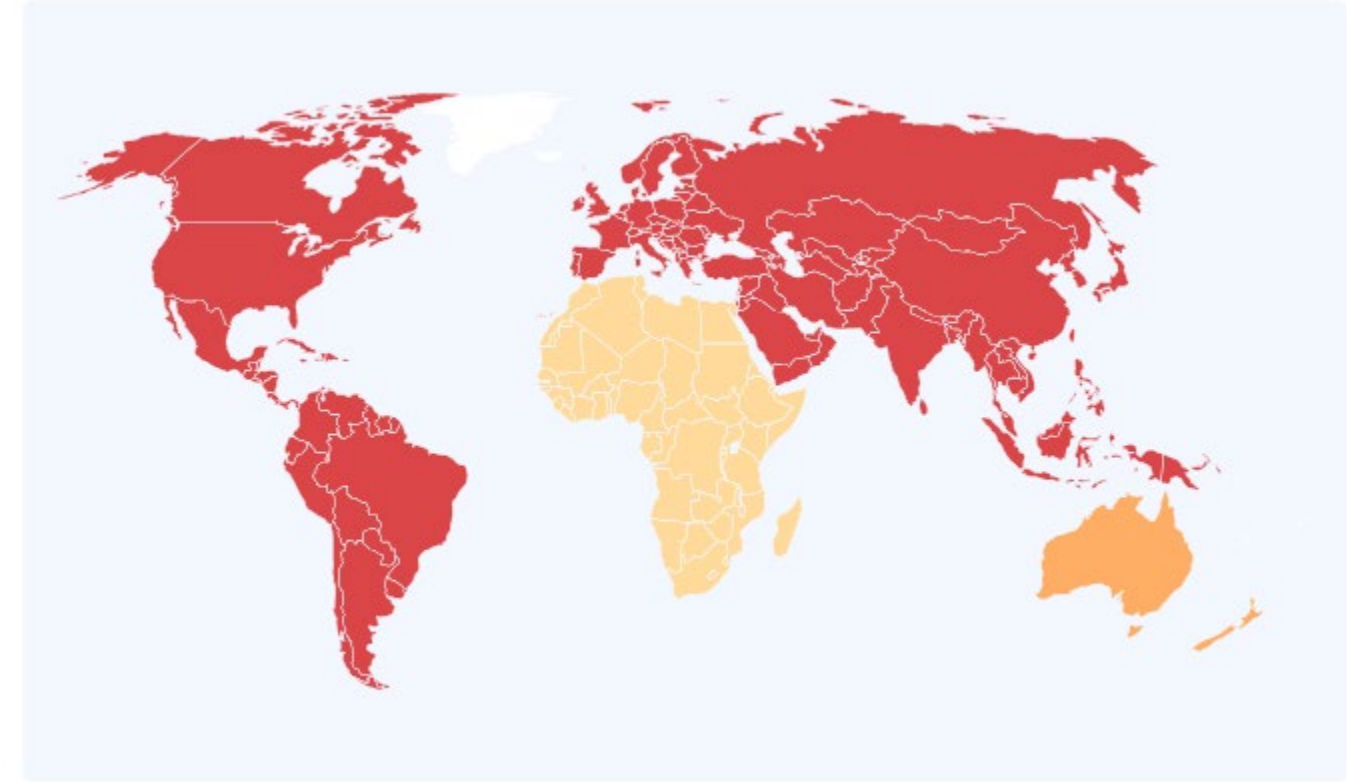
Current Status of Deepfakes



© 2024 Association of Certified Fraud Examiners, Inc.

Current Status of Deepfakes

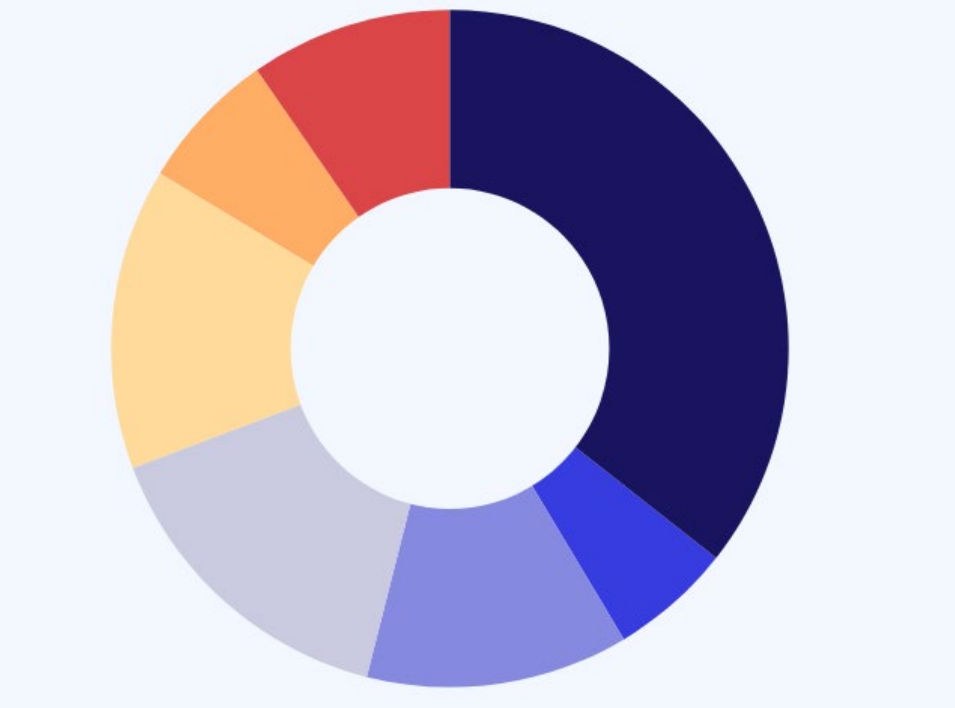
- Scam distribution map

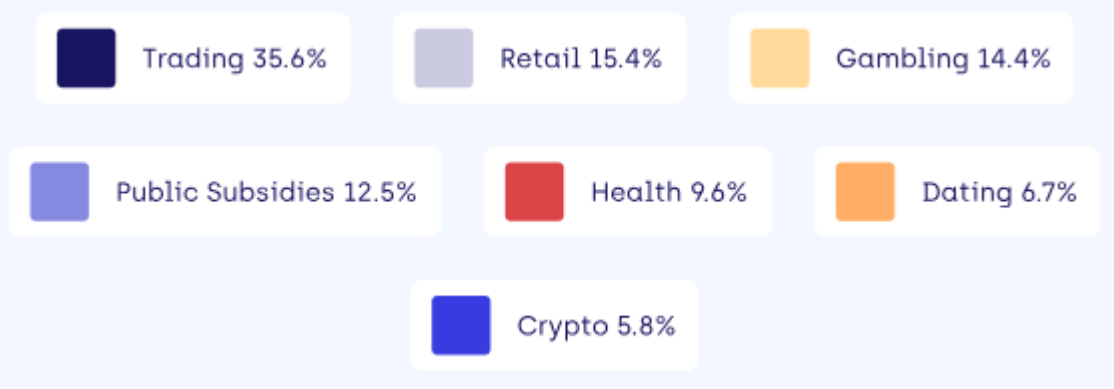


© 2024 Association of Certified Fraud Examiners, Inc.

Current Status of Deepfakes

- Most targeted industries





© 2024 Association of Certified Fraud Examiners, Inc.

Law Enforcement Warnings

- Europol report in 2022 on deepfake threats cites potential uses:
  - Harassment
  - Extortion and fraud
  - Facilitating document fraud
  - Falsifying online identities
  - KYC bypass
  - Falsifying or manipulating ditital evidence
  - Disrupting financial markets
- FBI warning in June 2022 on remote job application scams
- FBI warning in 2023 about increased use of deepfakes in “sextortion”
- FinCEN warning in November 2024 about generative AI fraud schemes targeting financial institutions

© 2024 Association of Certified Fraud Examiners, Inc.

FinCEN Generative AI Warning

- Increase in suspicious activity reporting by financial institutions in 2023 and 2024
- Alteration or creation of fraudulent identity documents to circumvent KYC protocols
  - Documents, photographs, and videos
  - Synthetic identities
- Specifically mentions open-sourcing of tools
- Opening accounts to receive and launder fraud proceeds
- Use of deepfake media in phishing attacks and scams
- Recommendations include multifactor authentication, live verification checks

© 2024 Association of Certified Fraud Examiners, Inc.

Deepfake Scam Applications

- BEC scams
- Pump and dump
- Market manipulation
- Faked endorsements
- Bribery and corruption
- Telecom frauds
- Identity theft
- Account takeovers
- SIM swap
- Telemedicine fraud
- Ransomware/extortion
- Virtual kidnapping
- Network penetration
- Data breaches
- Impersonation/imposter scams
- Synthetic IDs

© 2024 Association of Certified Fraud Examiners, Inc.

LLM Fraud Applications

- Convincing text for social engineering attempts
- Generation of fraudulent documentation
- Aids generation of fictitious companies
- Fake ID generation
- Automation of tasks

© 2024 Association of Certified Fraud Examiners, Inc.

Deepfake scams escalate, hitting 53% of

CEO of world's bigg
by deepfake scam

"Deepfake attacks now occur every five minutes,
"Entrust report warns

Report

users

North Kore
nuclear we

to create malware, says OpenAI;
Here's how

Hong Kong police bust fraud ring that used face-swapping tech for romance scams

© 2024 Association of Certified Fraud Examiners, Inc.

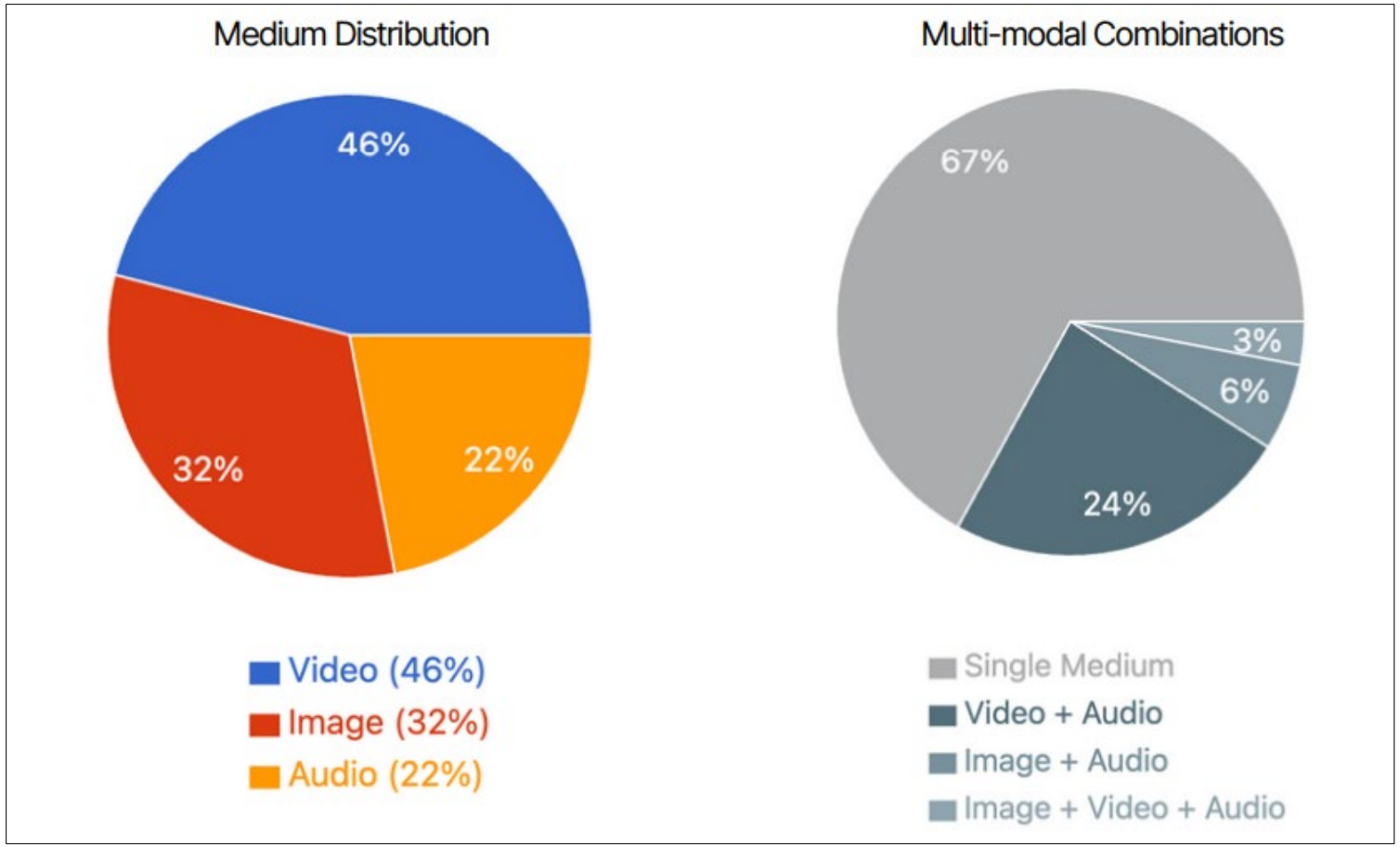
Entrust Identity Fraud Report

- Digital manipulation surpassed physical document manipulation for the first time in 2024
- Digital forgeries increased 224% year over year
- Deepfakes now account for 40% of document fraud

© 2024 Association of Certified Fraud Examiners, Inc.

Resemble AI Report

- Report from AI company who offers both voice generation and deepfake detection claimed they identified in Q1 2025
  - 163 documented deepfake incidents analyzed
  - More than $200 million lost in frauds involving deepfakes



© 2024 Association of Certified Fraud Examiners, Inc.

Business Email Compromise (BEC)

- Form of spear phishing in which fraudsters typically impersonate executive or vendor/supplier over email
- Typically involves wire transfers
- BEC facts according to FBI's IC3
  - 21,442 complaints in 2024 (actually down by a few dozen)
  - Adjusted losses $2.8 Billion (actually down about $100K)
  - More than $58 billion in losses reported to IC3 since October 2013

© 2024 Association of Certified Fraud Examiners, Inc.

Deepfake BEC

- In July 2019, cybersecurity company Symantec claimed it had verified three cases of deepfake audio in BEC scams that worked.
- In August 2019, insurer Euler Hermes Group disclosed a March 2019 deepfake BEC that impersonated German manufacturing CEO and cost the victim $243,000.
- Nikkei $29 million deepfake BEC is unconfirmed.
- U.A.E. company targeted with deepfake audio BEC in 2020, losing $35 million

© 2024 Association of Certified Fraud Examiners, Inc.

Deepfake BEC

- In July 2020, software company Nisos reported a deepfake audio attempt against a tech client that occurred in June 2020 but was unsuccessful.
  - Voice mail impersonating executive
  - Recipient found recording suspicious and immediately reported it to the organization’s legal department

© 2024 Association of Certified Fraud Examiners, Inc.

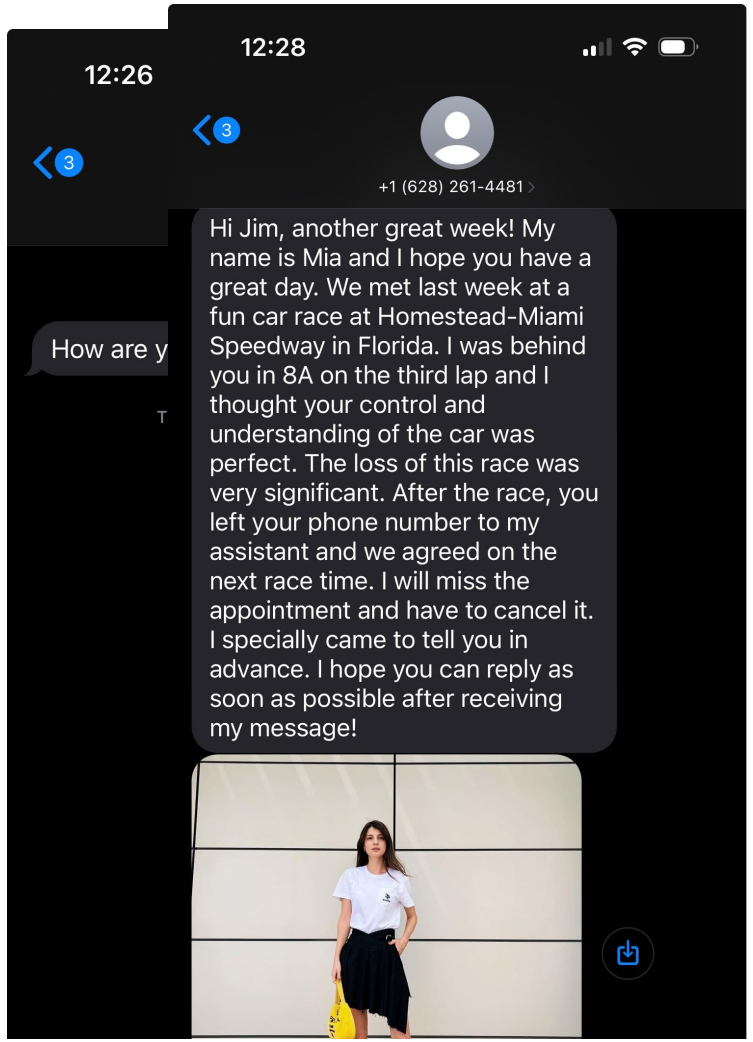
Deepfake BEC

- Finance worker at multinational firm tricked into paying out $25 million
- An initial message from the CFO asking for a “secret transaction” raised the employee’s suspicions, so he set up a video call
- The video call featured numerous colleagues, not just the requestor, so the victim felt comfortable making the transfer
- So, the fraudster was able to use multiple deepfake videos in a live setting to convince the victim
- This marks a significant escalation in sophistication of tactics

© 2024 Association of Certified Fraud Examiners, Inc.

Pig Butchering Scams

- Evolution of Romance Fraud that involves cryptocurrency investment
- Enhanced by Generative AI
- More likely to affect you personally than professionally
- Typically starts with apparent “wrong number” text or a message
- Billions in losses reported over the previous few years, with many unreported losses suspected



© 2024 Association of Certified Fraud Examiners, Inc.

Biometric Spoofing

- Biometric authentication use is increasing in anti-fraud controls:
  - Facial recognition
  - Voice recognition
- Most existing solutions are not completely accurate.
- Academic study from 2019 found that GAN-created videos fooled facial recognition systems in up to 95% of attempts.
- Attackers are reportedly bypassing KYC biometrics using “digital injection attacks”
- In Feb 2023, a Vice journalist used ElevenLabs’ free audio deepfake service to bypass bank biometrics and access his account

© 2024 Association of Certified Fraud Examiners, Inc.

Investigations and Evidence

- Can “beyond a reasonable doubt” exist anymore?
  - Prioritize other evidence?
  - Criminal versus civil?
- How do you authenticate audio/video evidence?
  - Updated chain of custody procedures?
  - Digital forensics?
- How do you record interviews?
  - Multiple formats?
  - Multiple devices?
  - Watermarks?

© 2024 Association of Certified Fraud Examiners, Inc.

Investigations and Evidence

- Are you using ChatGPT or other LLMs for investigation?
  - Data analysis – probably ok
  - Outsourcing your investigation and/or critical thinking skills – not ok
- The more you control the input on LLM prompts, the less of a risk of “hallucinations”
- Several legal professionals have gotten in trouble for using LLMs for court docs
- Proposed changes to rules of evidence include giving more power to judges to restrict audio/video evidence, force parties who claim evidence is deepfake to pay for forensic analysis, but would take years to go into effect

© 2024 Association of Certified Fraud Examiners, Inc.

Quick Case Study #1

- Teen cheerleader's mother accused of using deepfakes to defame competitors
- Mother arrested in March 2021 for allegedly manipulating images and videos of inappropriate behavior
- Charges of cyber harassment of a child specifically mentioned "deepfake"
- No evidence provided to substantiate claims, no original video to analyze, researchers skeptical

© 2024 Association of Certified Fraud Examiners, Inc.

Quick Case Study #2

- Baltimore school principal suspended due to audio featuring racist remarks about school children being circulated on social media
- Months later, the school's former athletic director was arrested and charged with disrupting school activities for creating the audio clip using AI and circulating it
- The school system had to arrange for police presence at the principal's house during the investigation
- The former athletic director was found to have searched for AI tools using school devices, and was one of three educators who reportedly received the recording the night before it was widely circulated.

© 2024 Association of Certified Fraud Examiners, Inc.

Technical Solutions

- Numerous deepfake detection or video authentication providers dedicated to deepfakes (Authme, Sensity, Pindrop, deepware, Thales, Surf Security iProov, etc.)
- Institutional companies offering solutions (Intel FakeCatcher, Microsoft Video Authenticator, DeepMind SynthID)
- Blockchain platforms?
- Watermarked videos?
  - Deepfake generator app Avatarify said it would include digital watermark, but you can pay a premium to have it removed
  - castLabs watermarking tool

© 2024 Association of Certified Fraud Examiners, Inc.

Technical Solutions

- Accuracy?
- Practicality?
- Cost?
- Applicability to need?
- Futility? There's no stopping synthetic content generation on the internet at this point

© 2024 Association of Certified Fraud Examiners, Inc.

Potential Controls

- Multi-factor **authorizations**
- Independent verification via trusted channel
- Hang up and call back
- Passphrases
- Block VOIP calls and/or known malicious IP addresses
- No more voice recognition
- Behavioral biometrics
- Update IT and wire transfer protocols

© 2024 Association of Certified Fraud Examiners, Inc.

Why is This Important for Fraud Examiners?

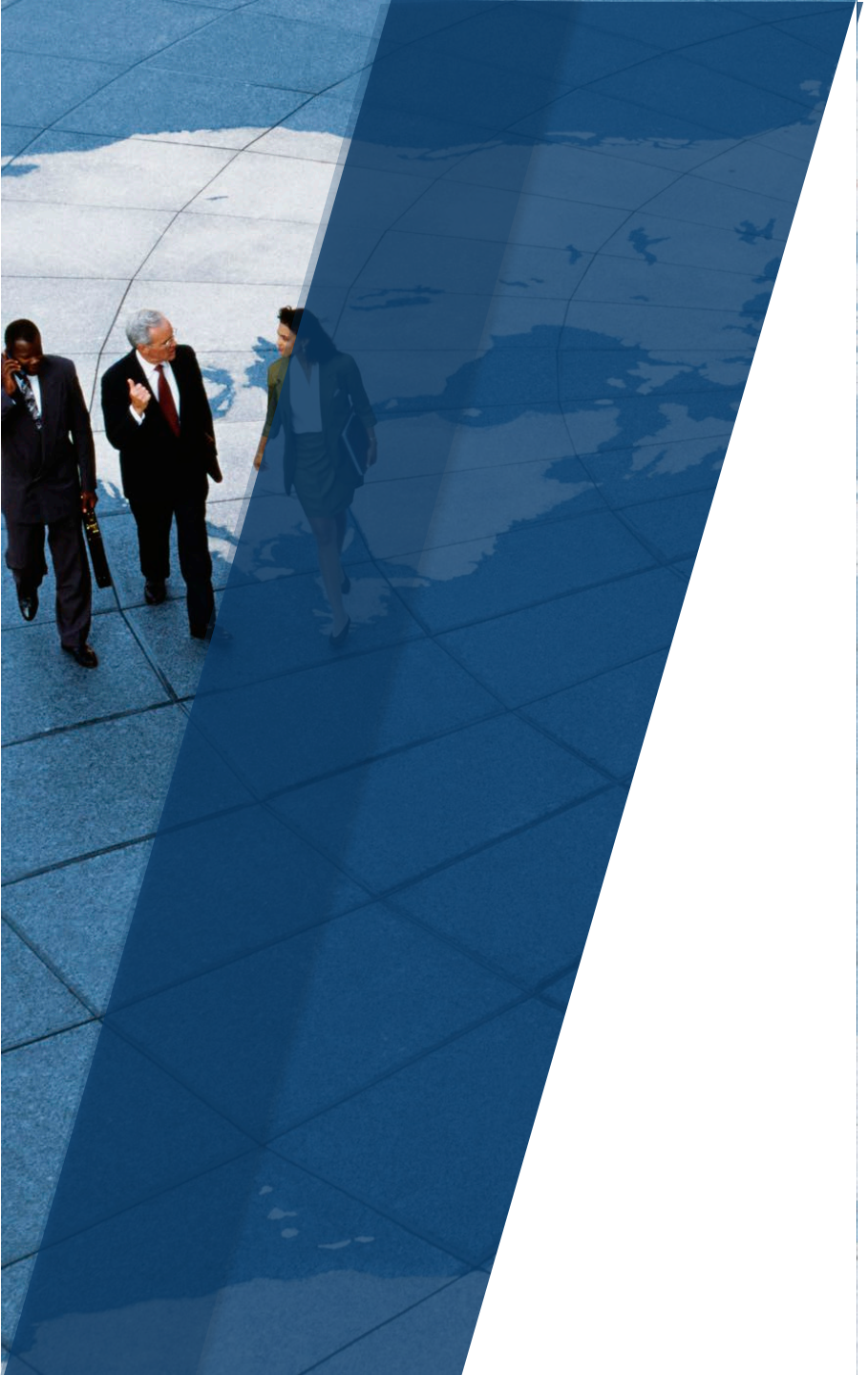
- These fraud risks are real and increasing
- Awareness and training opportunities
- **Keep corroboration and chain of custody in mind throughout an investigation**
- Look at technical solutions for identification of AI-generated content as they emerge, but be skeptical
- Consider updating skill set (forensics?) or bringing in experts
- **Do not rely on information provided by ChatGPT/Gemini/etc.**
- Consider applications of Generative AI for investigative/reporting tasks but be careful

© 2024 Association of Certified Fraud Examiners, Inc.

In Summary

- It’s already too late to stop the threats, which are many.
- Spread awareness in your organization and your personal networks
- Technology gets better and more accessible, and there’s already zero barrier for entry.
- Technical solutions will involve relying on unreliable platforms and always be playing catch-up.
- Use with caution, consider liabilities.
- Hone your skills.

© 2024 Association of Certified Fraud Examiners, Inc.



Mason Wilder, CFE
ACFE Research Director
mwilder@acfe.com