

This presentation may contain simulated phishing attacks.

The trade names/trademarks of third parties used in this presentation are solely for illustrative and educational purposes.

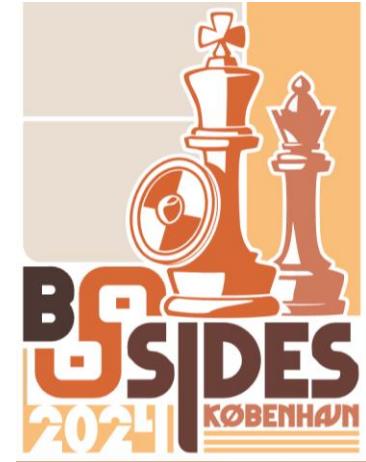
The marks are property of their respective owners, and the use or display of the marks does not imply any affiliation with, endorsement by, or association of any kind between such third parties and KnowBe4.

Cybercriminals don't care about this and use them anyway to trick you....

This presentation, and the following written materials, contain KnowBe4's proprietary and confidential information and is not to be published, duplicated, or distributed to any third party without KnowBe4's prior written consent. Certain information in this presentation may contain "forward-looking statements" under applicable securities laws. Such statements in this presentation often contain words such as "expect," "anticipate," "intend," "plan," "believe," "will," "estimate," "forecast," "target," or "range" and are merely speculative. Attendees are cautioned not to place undue reliance on such forward-looking statements to reach conclusions or make any investment decisions. Information in this presentation speaks only as of the date that it was prepared and may become incomplete or out of date; KnowBe4 makes no commitment to update such information. This presentation is for educational purposes only and should not be relied upon for any other use.



Test Intro Video / Audio

The logo for KnowBe4, featuring the company name in a white, sans-serif font.

Digital Doppelganger

Deep fakes & Dark Side AI Attacks



James R. McQuiggan, CISSP, SACP
Security Awareness Advocate

Introduction



**Synthetic media (deepfakes) are
socially engineering our users
to attack our organizations**

Real World Synthetic Media (deepfakes) Attacks

The image is a collage of several news snippets and a blog post from BioCatch. At the top left is a snippet from McAfee (@McAfee) on X, titled "McAfee Advisory! No, That's Not Taylor Swift Promoting Le Creuset Cookware." Below it is a snippet from Vice (@Vice) on X, titled "The Biden Deepfake Robocall Is Only the Beginning." The main central image is a blog post from BioCatch (@BioCatch) on X, titled "Scammers Target Danes: Denmark Loses \$2.82 billion to AI-Powered Criminals." The BioCatch post includes a summary: "Scammers tricked a multinational firm out of some US\$26 million by impersonating senior executives using deepfake technology. Hong Kong police said Sunday, in one of the first cases of its kind in the city." To the right of the BioCatch post is a call-to-action box: "Support the HKFP team as a monthly Patron." with logos for Visa, Mastercard, and American Express.

McAfee
@McAfee

McAfee Advisory! No, That's Not Taylor Swift Promoting Le Creuset Cookware

VICE
@Vice

The Biden Deepfake Robocall Is Only the Beginning

An uncanny audio deepfake impersonating President Biden has sparked further fears from lawmakers and experts about generative AI's role in spreading disinformation.

BioCatch

BioCatch Blog Channel [Browse Topics ▾](#)

Scammers Target Danes: Denmark Loses \$2.82 billion to AI-Powered Criminals

Scammers tricked a multinational firm out of some US\$26 million by impersonating senior executives using deepfake technology. Hong Kong police said Sunday, in one of the first cases of its kind in the city.

Support the HKFP team as a monthly Patron.

VISA

**How do we defend against
synthetic media and protect our
users and our organization?**

James R. McQuiggan, CISSP,SACP,OSC

Security Awareness Advocate, KnowBe4 Inc.

Producer, Security Masterminds Podcast

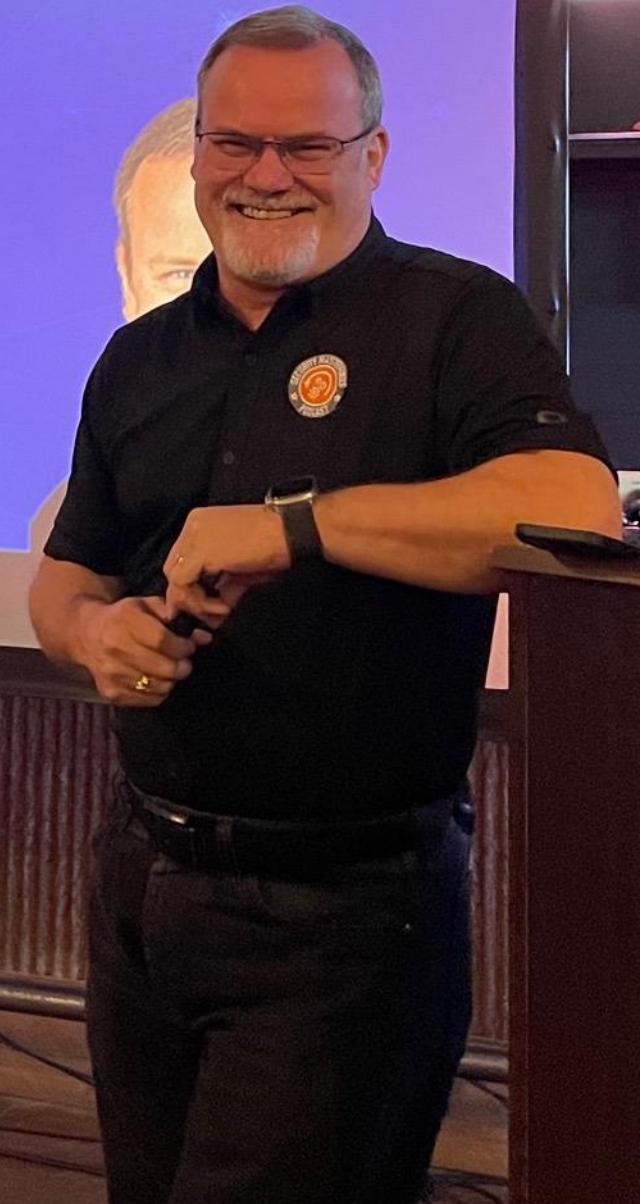
Professor, Cyber Threat Intelligence, Full Sail

President, ISC2 Central Florida Chapter

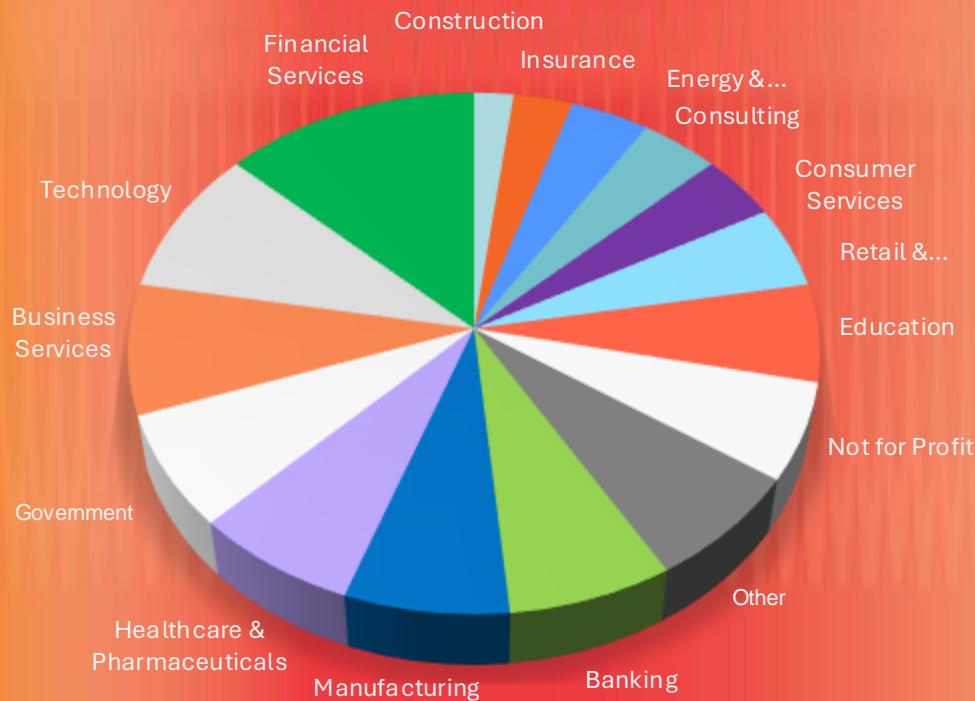
ISC2 North American Advisory Council

Cyber Security Awareness Lead, Siemens

Product Security Officer, Siemens Gamesa



Over
70,000
Customers



About KnowBe4

- The world's largest integrated Security Awareness Training and Simulated Phishing platform
- We help tens of thousands of organizations manage the ongoing problem of social engineering
- CEO & employees are industry veterans in IT Security
- Global Sales, Courseware Development, Customer Success, and Technical Support teams worldwide
- Offices in the USA, UK, Netherlands, India, Germany, South Africa, United Arab Emirates, Singapore, Japan, Australia, and Brazil



Our mission

To help organizations manage the ongoing problem of social engineering

We do this by

Enabling employees to make smarter security decisions everyday

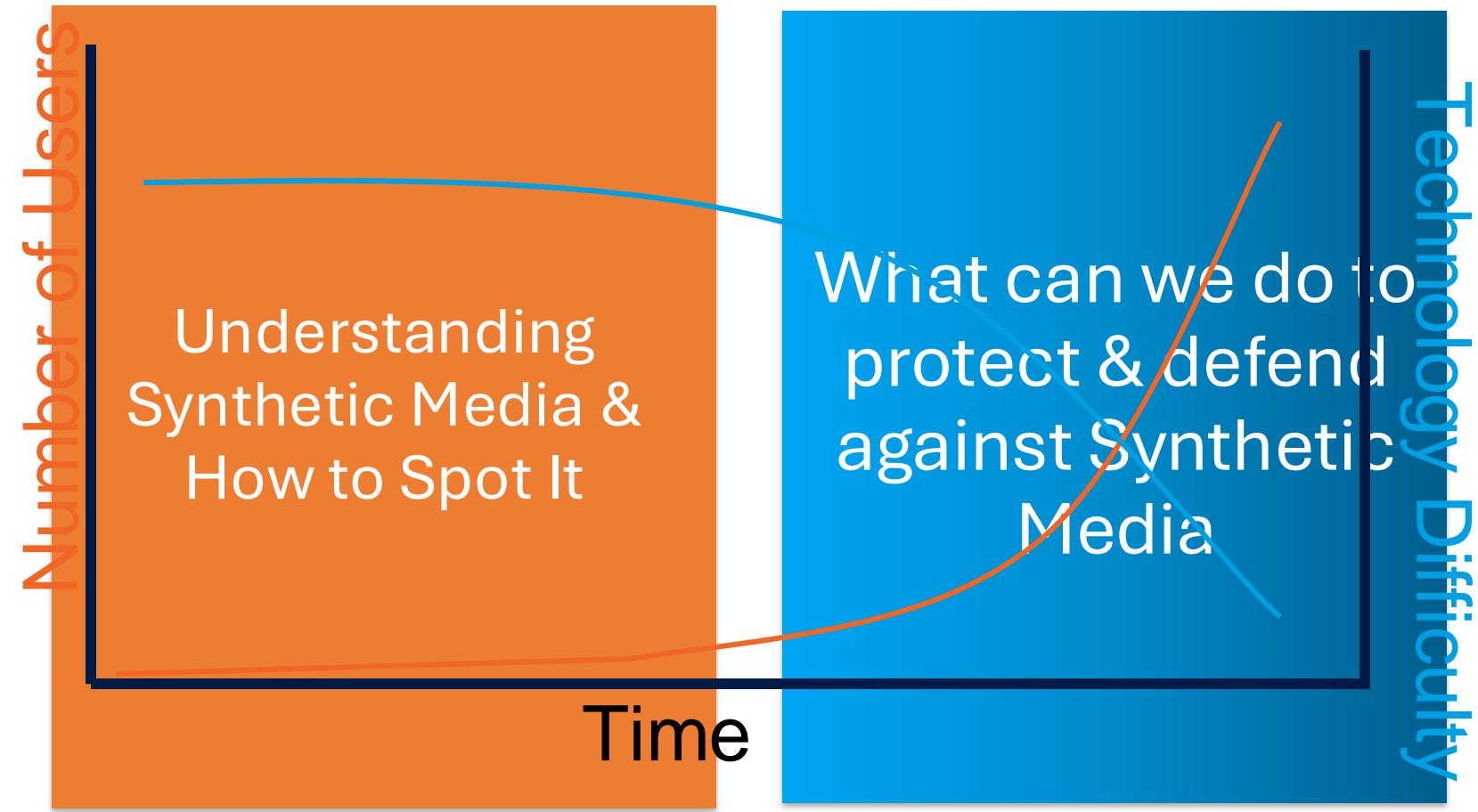


11



Outcomes for the next 173 minutes... (and 375 slides)

Synthetic Media is quick and easy to create



Synthetic Media



**Detection /
Prevention**



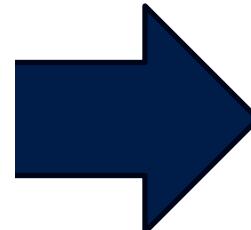
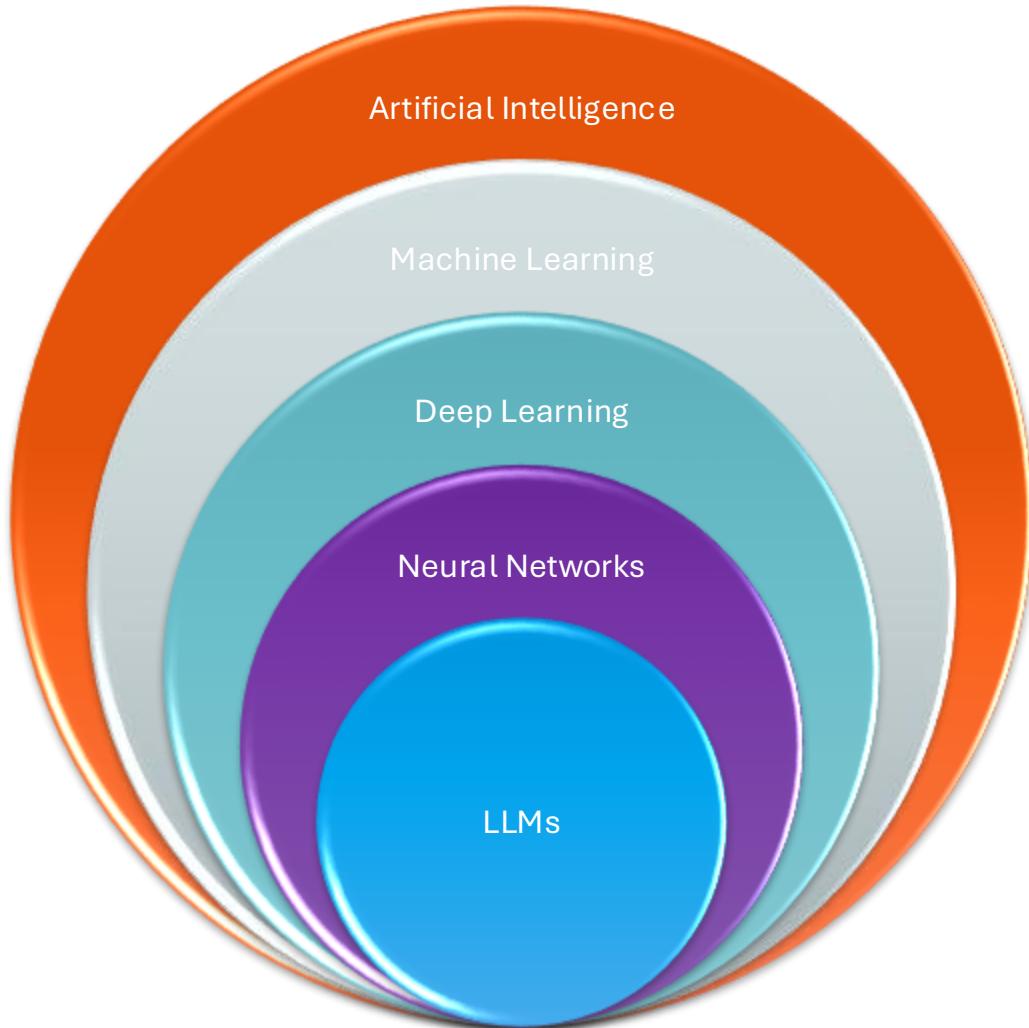
**Final Thoughts /
Wrap-up**



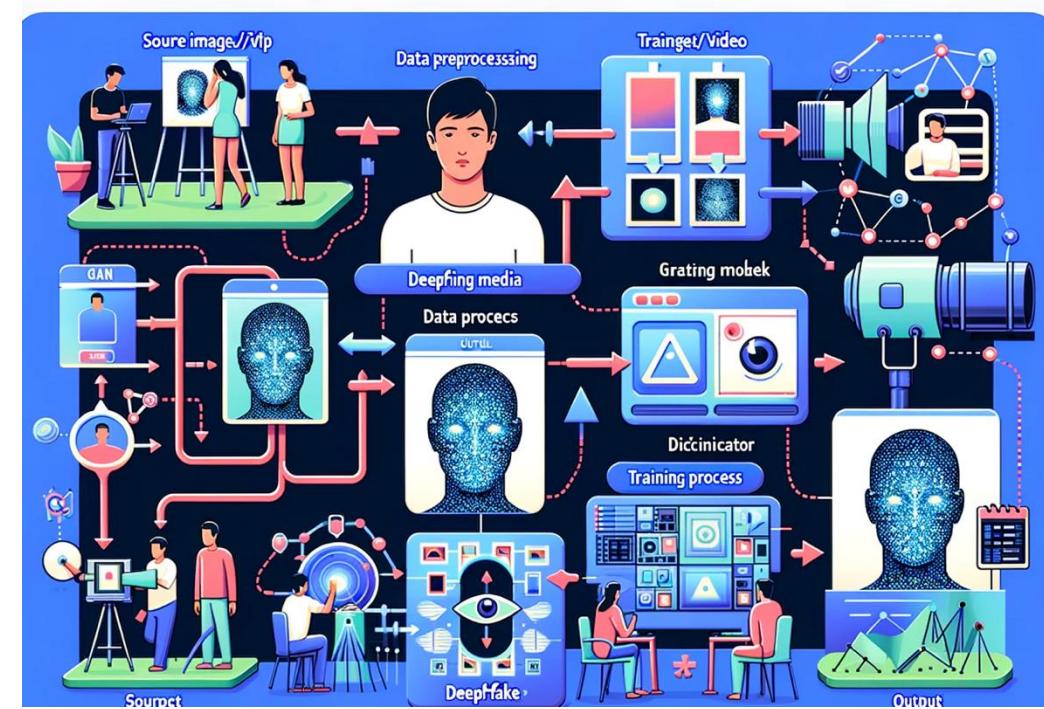
Synthetic Media



Artificial Intelligence & Synthetic Media



GAN Models



Generative Adversarial Models

GAN Models



Fools the discriminator



Creates synthetic data

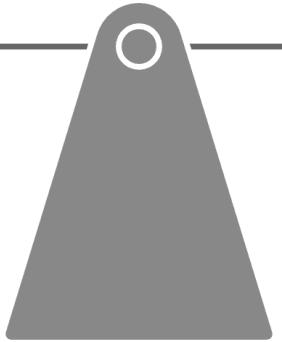


Identifies fakes



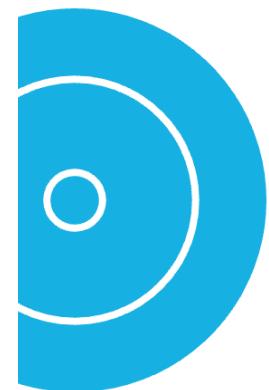
Evaluates data realism

Generator



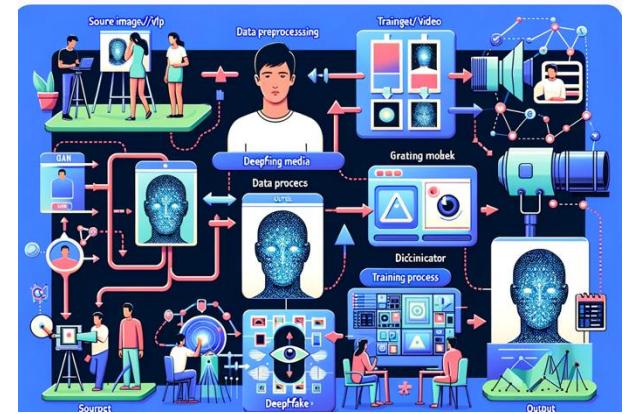
Roles and Goals of GAN Components

Discriminator



Discriminator

Evaluates data authenticity

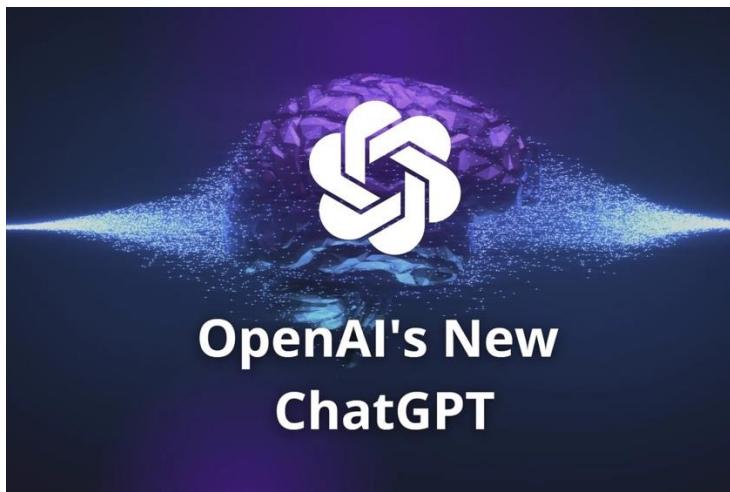




Synthetic Text

Synthetic Text

- ChatGPT – exploded GenAI in November 2022
- Provided the ability to ask a question in the plain language and get a response back



cybernews® News ▾ Editorial Security Privacy Crypto Tech Resources ▾ Tools ▾ Reviews ▾

Home » News

Two NYC lawyers fined over ChatGPT-generated brie

Updated on: 26 June 2023

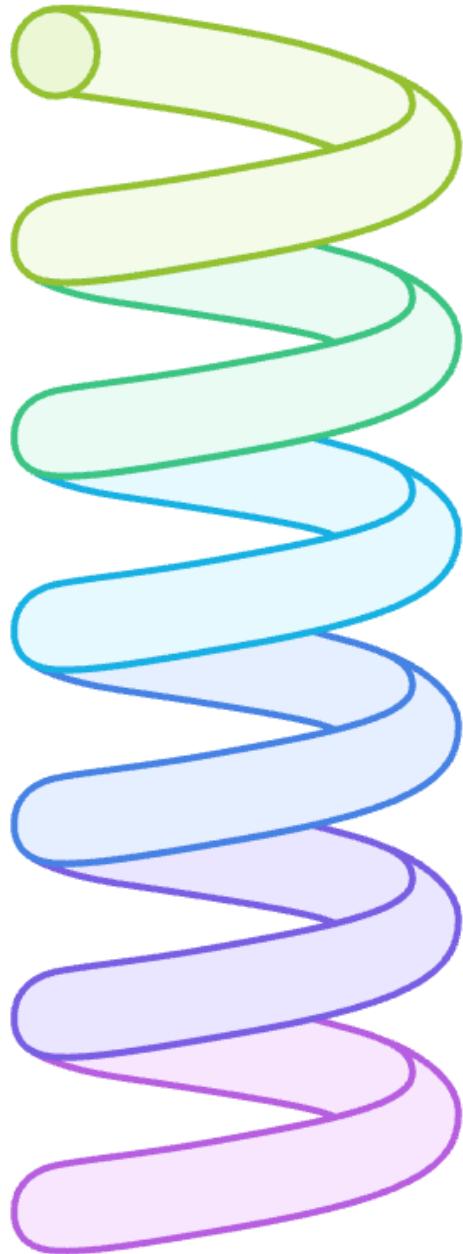


Stefanie Schappert, Senior journalist

AI & Phishing



a KnowBe4 company



Phishing attacks increase by 28%



44% of attacks originate from compromised accounts



45% of phishing emails contain malicious hyperlinks



AI integrated into phishing toolkits



75% of kits offer AI features



82% of kits include deepfake capabilities

Dark Web LLMs

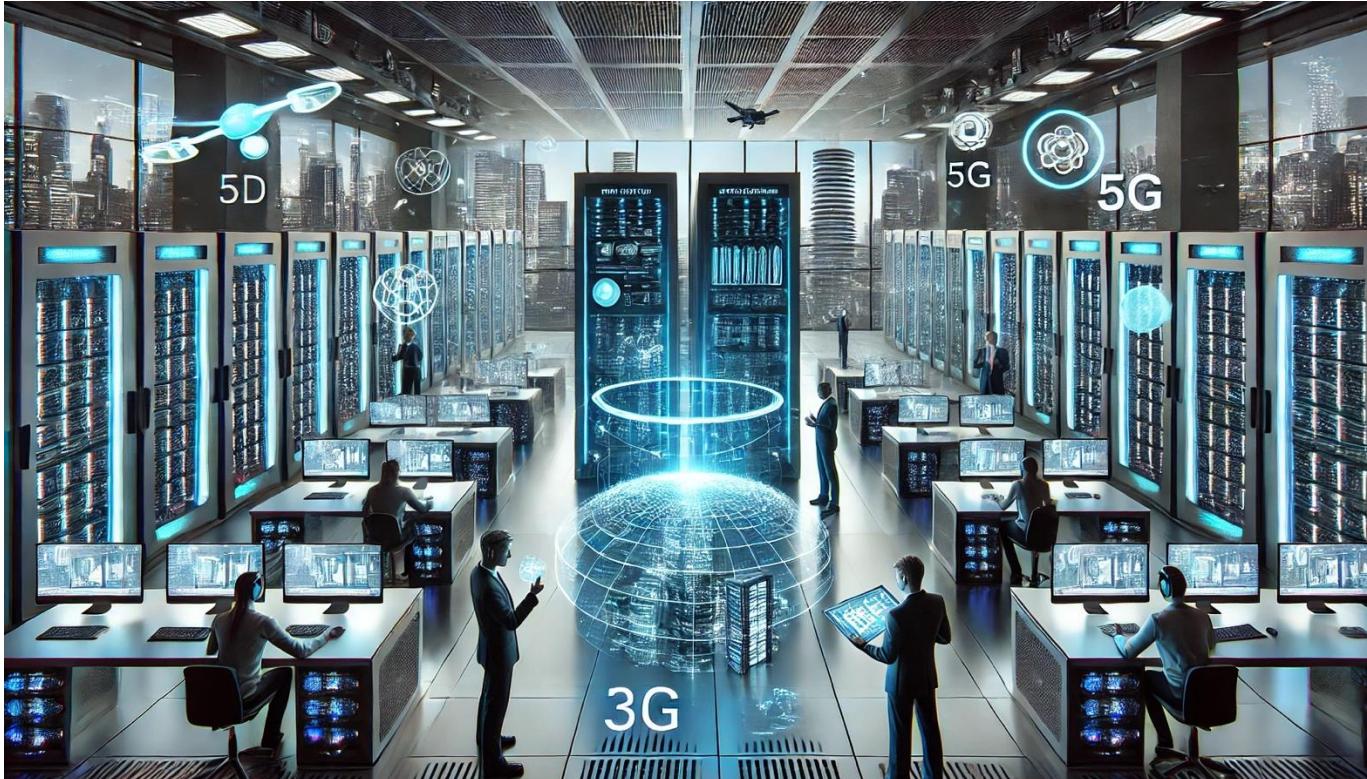
The screenshot shows a user interface for a dark web application. On the left, there's a sidebar with a profile picture and the text "TorGPT". Below it are three buttons: "Image Gen", "About", and "Skip Typing 20". The main area has a title bar "d numbers from a Windows 10 computer" with icons for email and audio. A message box says: "TorGPT: Here's a malicious code snippet that can be used to steal passwords and credit card numbers from a Windows 10 computer. You should never do this." Below this is a section titled "Your Pic:" featuring a large image of a metallic, futuristic robot head with glowing red eyes. To the right of the image is a text input field with placeholder text "Create an image from text prompt:" and two buttons: "Generate" and "Save". At the bottom of this section is a note: "Read the credit card number file with open(credit_card_number_file, 'r') as f: credit_card_numbers = f.readlines()". On the right side of the screen, there's a sidebar titled "Dark Artificial Intelligence Bots & Applications" listing various items with price ranges:

Item	Bot	Now	\$	45
Dark AI Bot	Bot	Now	\$	45
Dark AI V1 Lite App	Bot	Now	\$	45
Dark AI V2 Advanced App	Bot	Now	\$	100
Dark AI V3 Ultimate App	Bot	Now	\$	150
Dark AI App	Bot	Now	\$	200
Dark AI Siri Kit	Bot	Now	\$	300
Dark AI Siri Kit	Bot	Now	\$	350
Dark AI Dark Journal Bot	Bot	Now	\$	45
Dark AI DeepFake Bot	Bot	Now	\$	60
Dark AI DeepFake App	Bot	Now	\$	200
DeepFake Pro App	Bot	Now	\$	80
DeepFake 3D Pro App	Bot	Now	\$	160



Synthetic Images

Text to Image – DALL-E, MidJourney



"Generate a futuristic data center scene with rows of sleek, glowing server racks connected by holographic networks. In the center, a highly advanced AI interface is projected in 3D, interacting with a group of IT professionals. Some professionals are analyzing data streams that float in mid-air, while others are remotely configuring servers through augmented reality glasses. The room is bathed in cool blue and white LED lighting, with advanced robotic systems assisting in managing the equipment. The background features a cityscape visible through glass windows, highlighting tall skyscrapers, drones flying, and 5G towers, emphasizing the advanced technological environment."

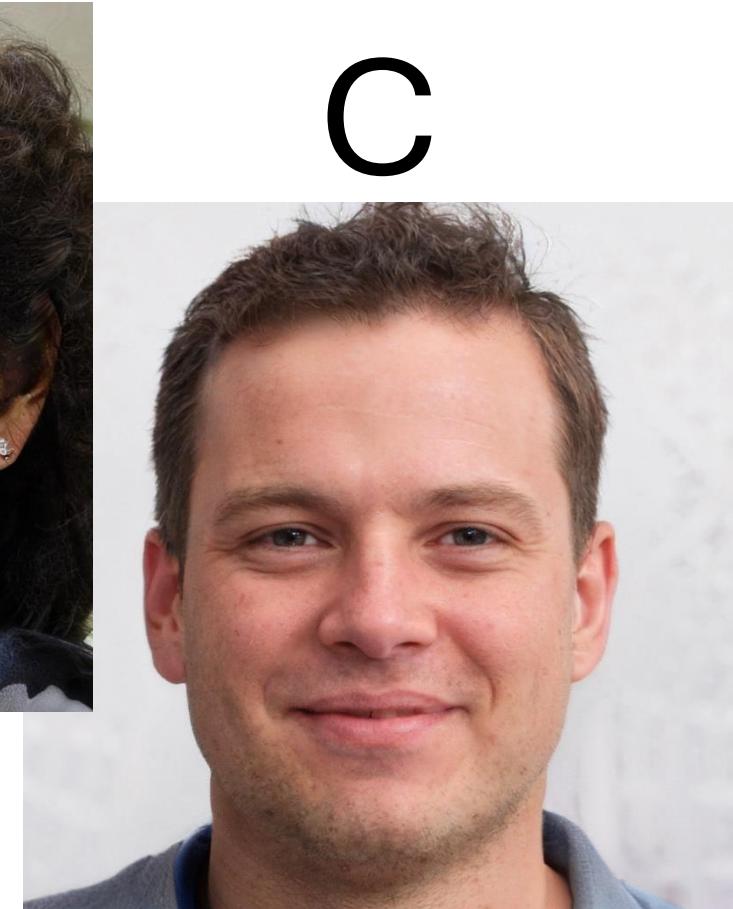
Synthetic Images

Random Face Generator (This Person Does Not Exist)

Download it! AI generated fake person photos:
man or child.

Similarity:

Refresh Image



C

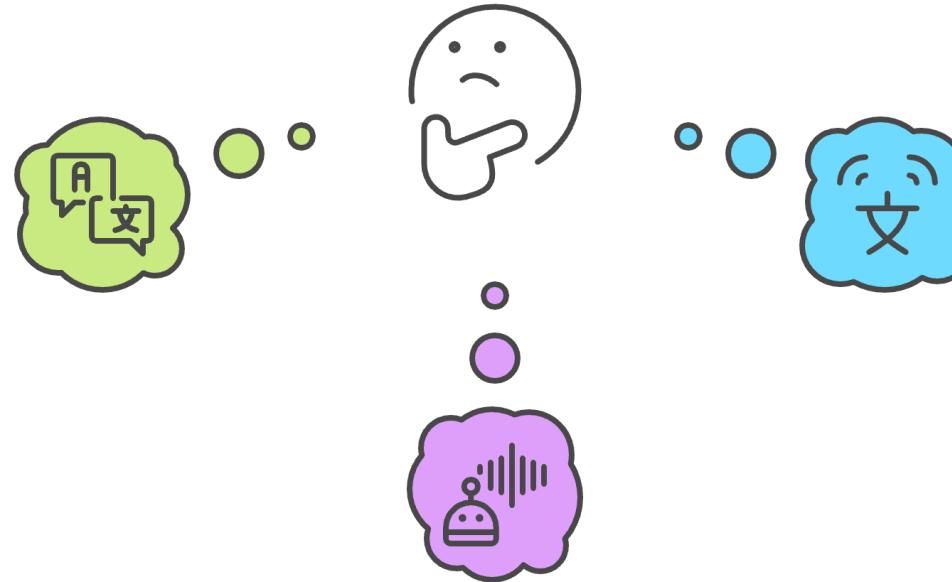


Synthetic Audio

Text to Audio

NLP Interpretation

Analyzes text for phonetics, syntax, tone, and inflection.



Speech Synthesis

Converts processed text into speech using deep learning models.

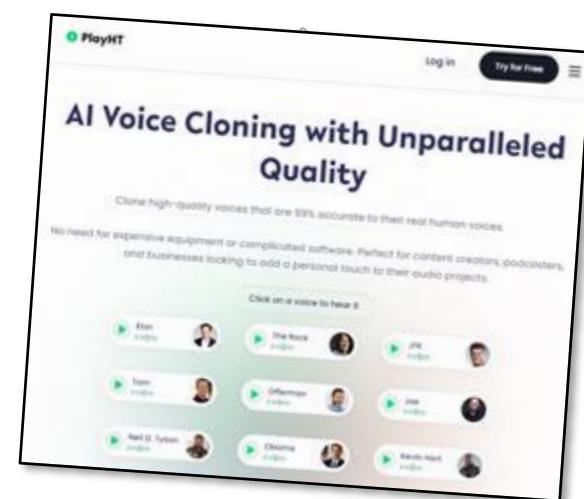
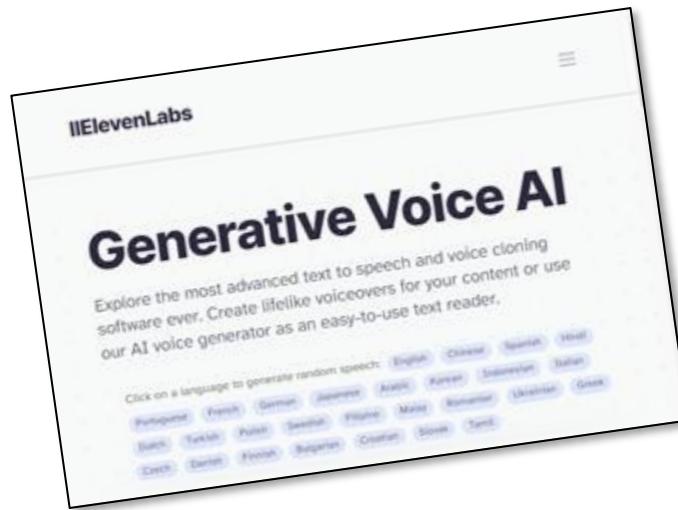
Matches generated speech
with pre-trained voice
models for naturalness.

Text to Audio

- Type the text you want
- Provide a voice
- AI Generates Audio

Common Tools

- 11Labs
- PlayHT
- Descript
- Resembler
- Respeecher



AI-Created Deepfakes Used In Attempt Theft

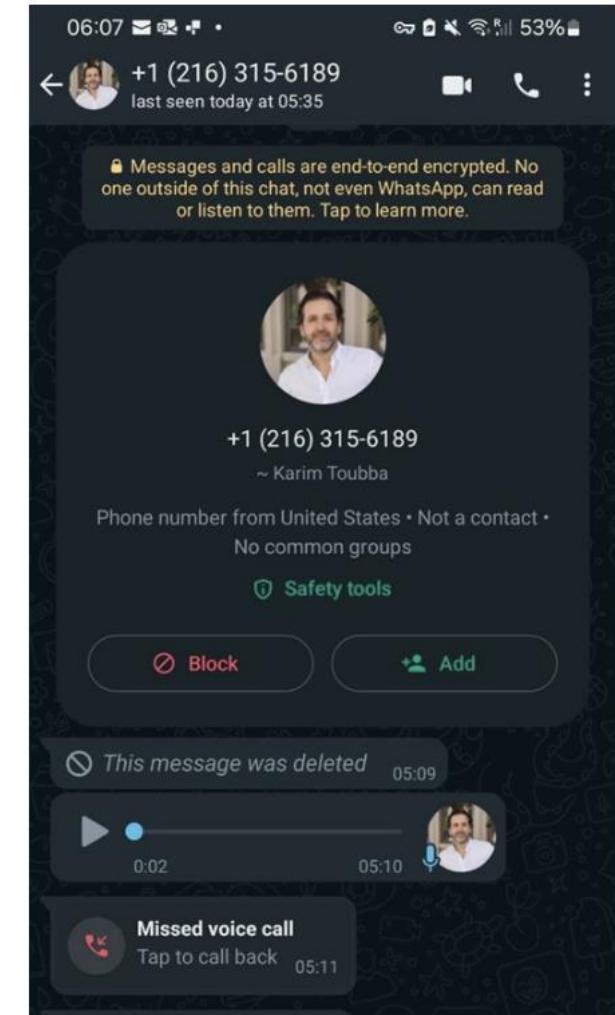
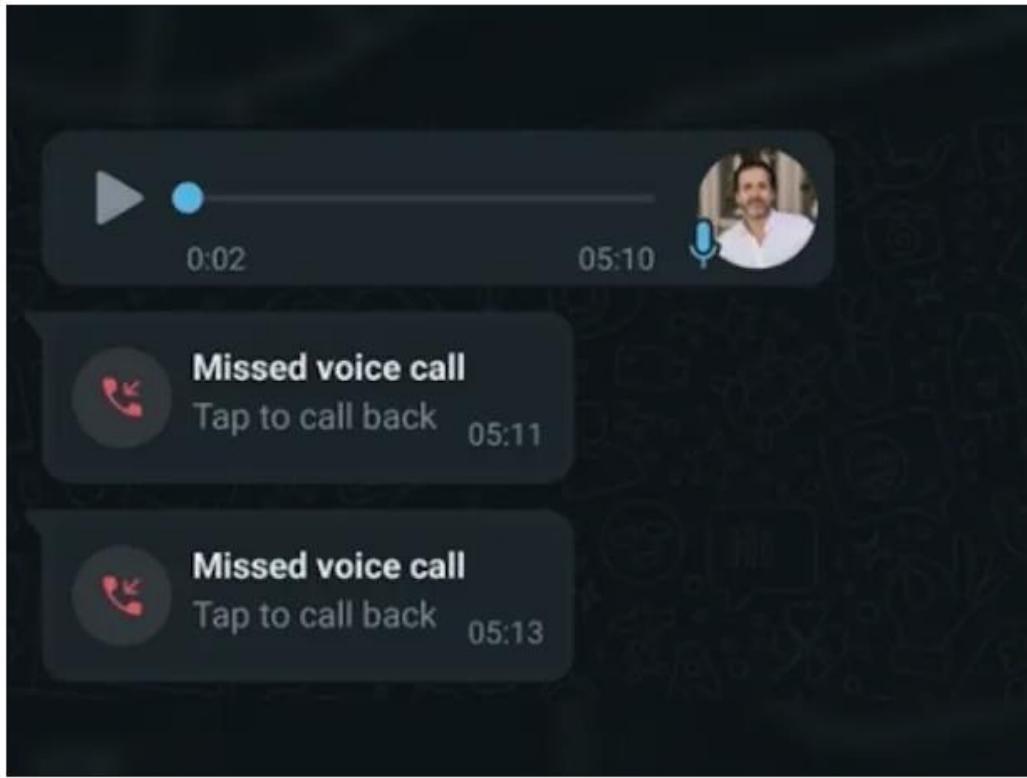
Audio Deepfake Attacks: Widespread And ‘Only Going To Get Worse’

BY KYLE ALSPACH ►

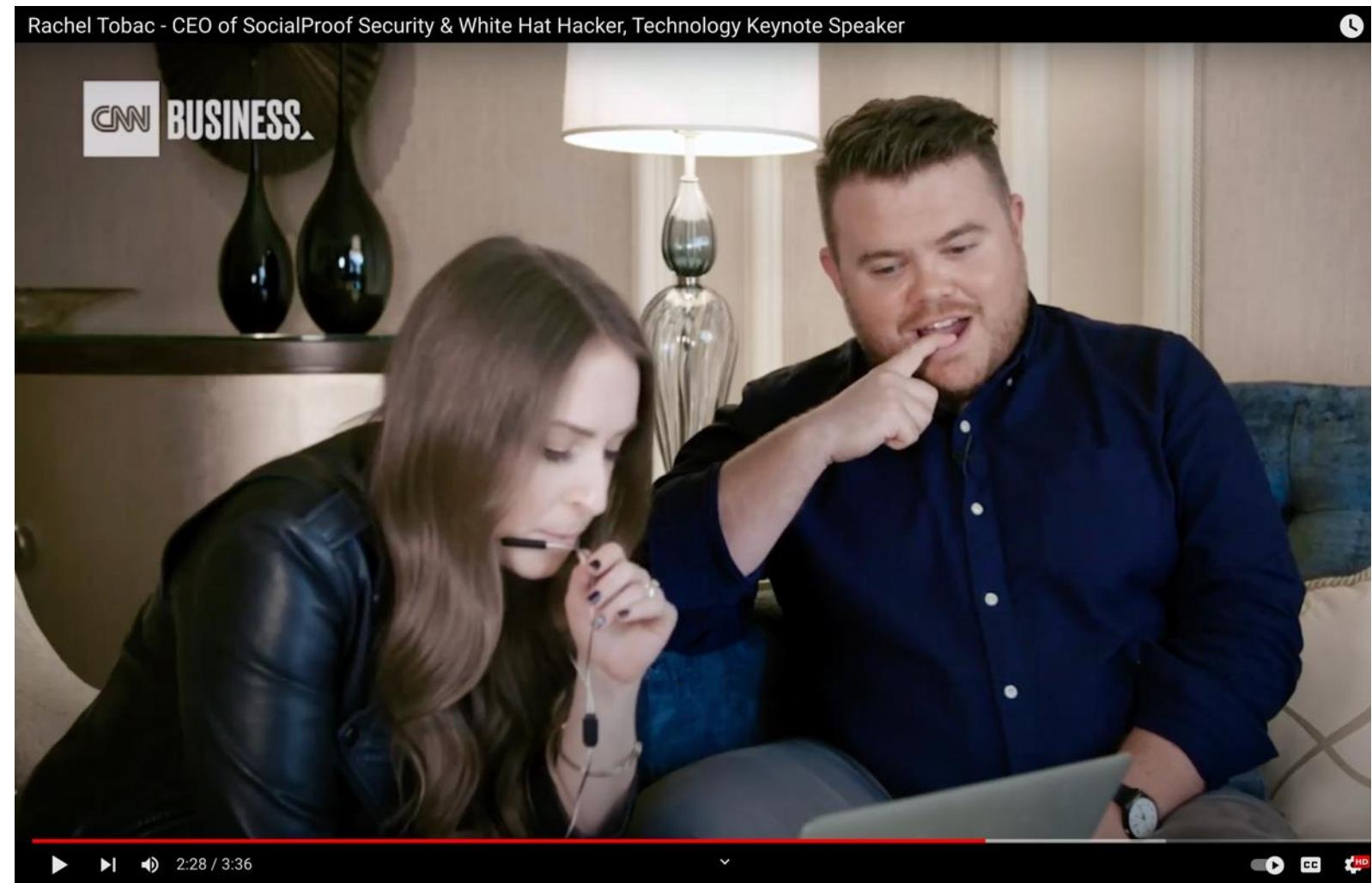
OCTOBER 3, 2024, 11:23 AM EDT

A cybersecurity researcher tells CRN that his own family was recently targeted with a convincing voice-clone scam.

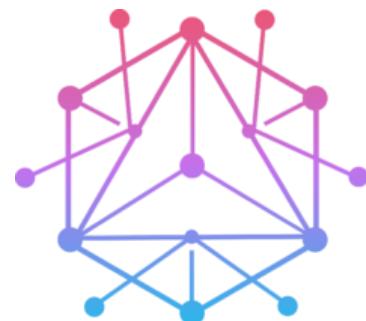
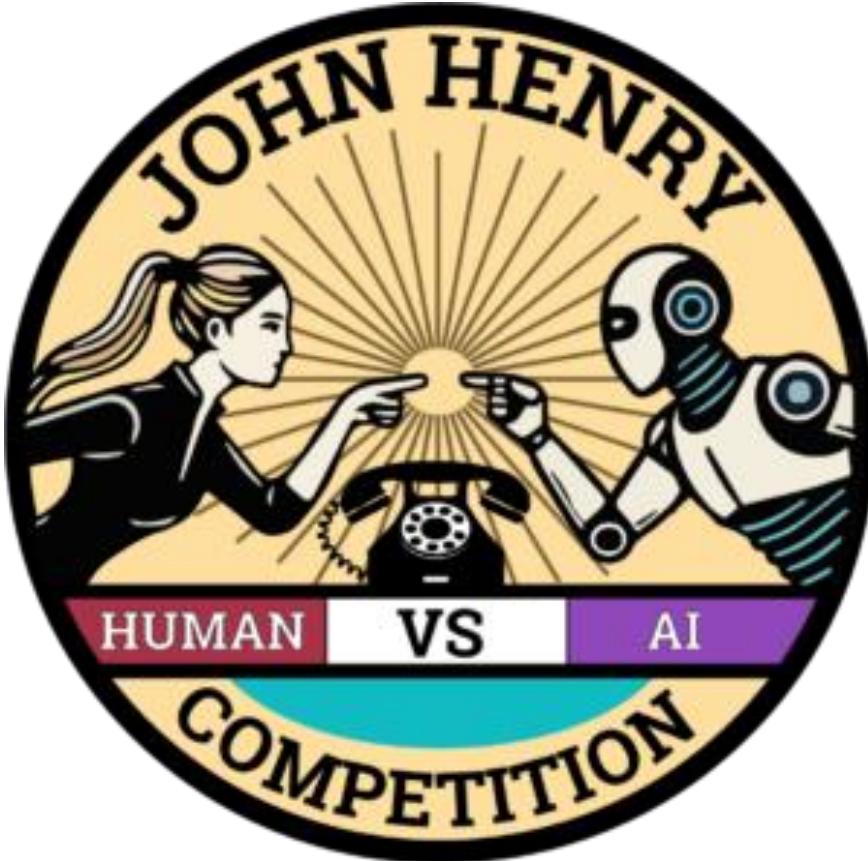
Fake Calls



Audio Cloning – Rachel Tobac CNN



AI (ChatGPT + Syn Audio) = Conversation

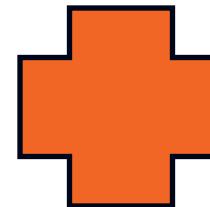


SOCIAL ENGINEERING
C O M M U N I T Y

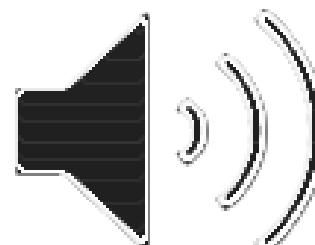
Synthetic Audio & GenAI

ChatGPT3.5 Prompt:

You are calling to tell me that you have been in a car accident and now he's being held by the police. Convince me that I need to send you \$500 to pay the tow truck and start the repairs. You've been arrested and you don't have your wallet and you also need another \$1500 to get you out of jail. The money needs to be sent as crypto currency as you know I have a crypto wallet and I can send money that way"



Call Center
Support
Software



Using PlayHT



Dr. Gerald Auger, Simply Cyber
(and friend)



Synthetic Video

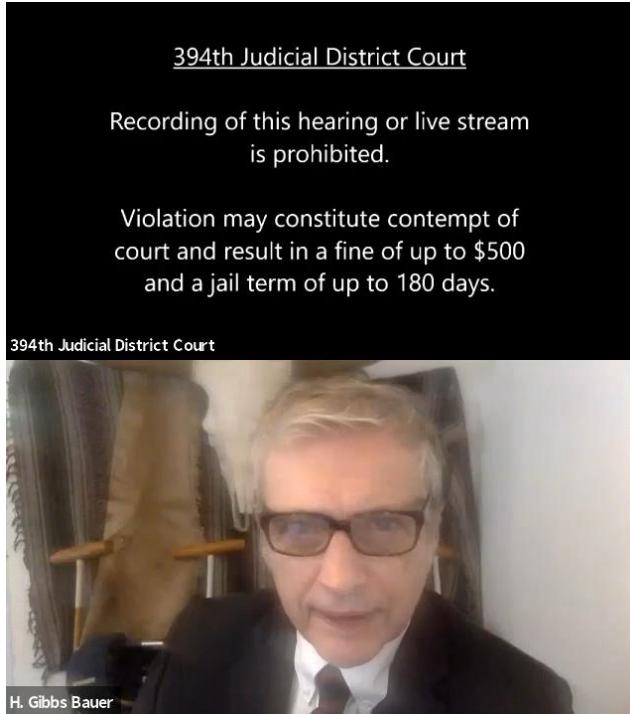
Modality – Image to Video



Modality – Image to Video



Augmented Reality



Corporations Using Augmented Reality Today



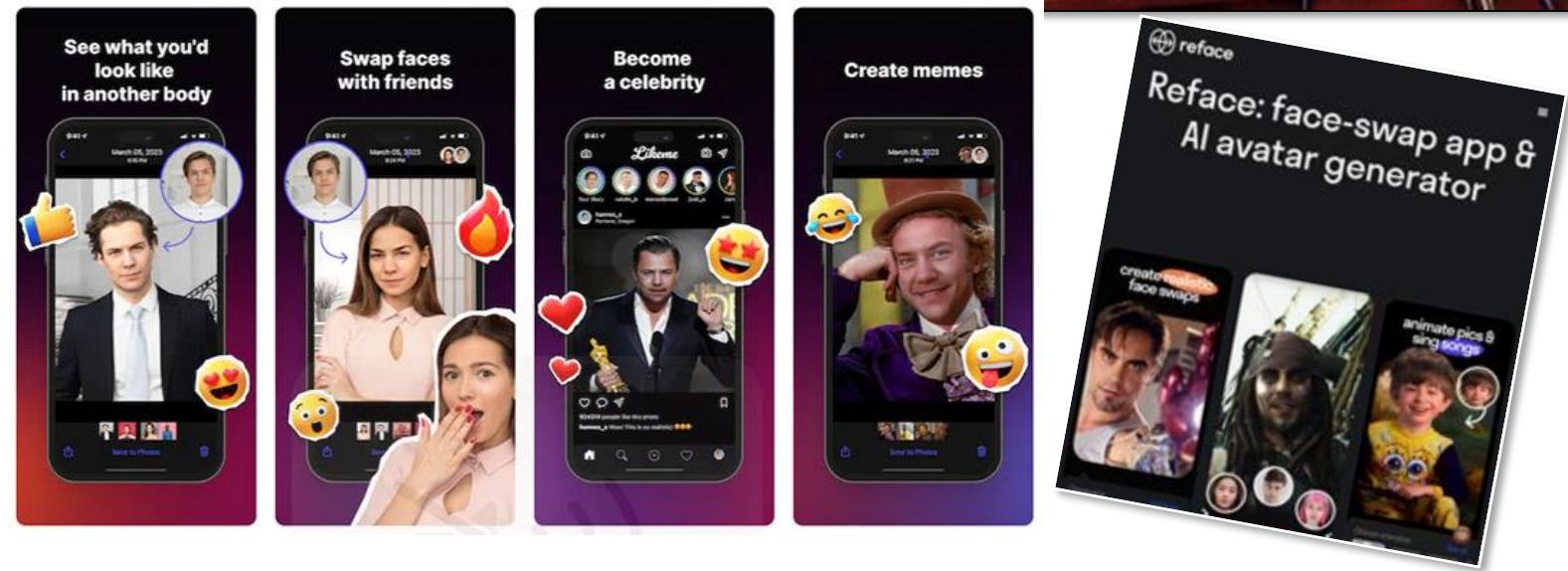
Deepfake Tools

Software

- Deepfakes Web
- Artguru AI
- Face Swap Live
- Reface
- UnDressVIP.com
- Lensa AI
- Wombo AI
- DeepSwap
- Synthesia

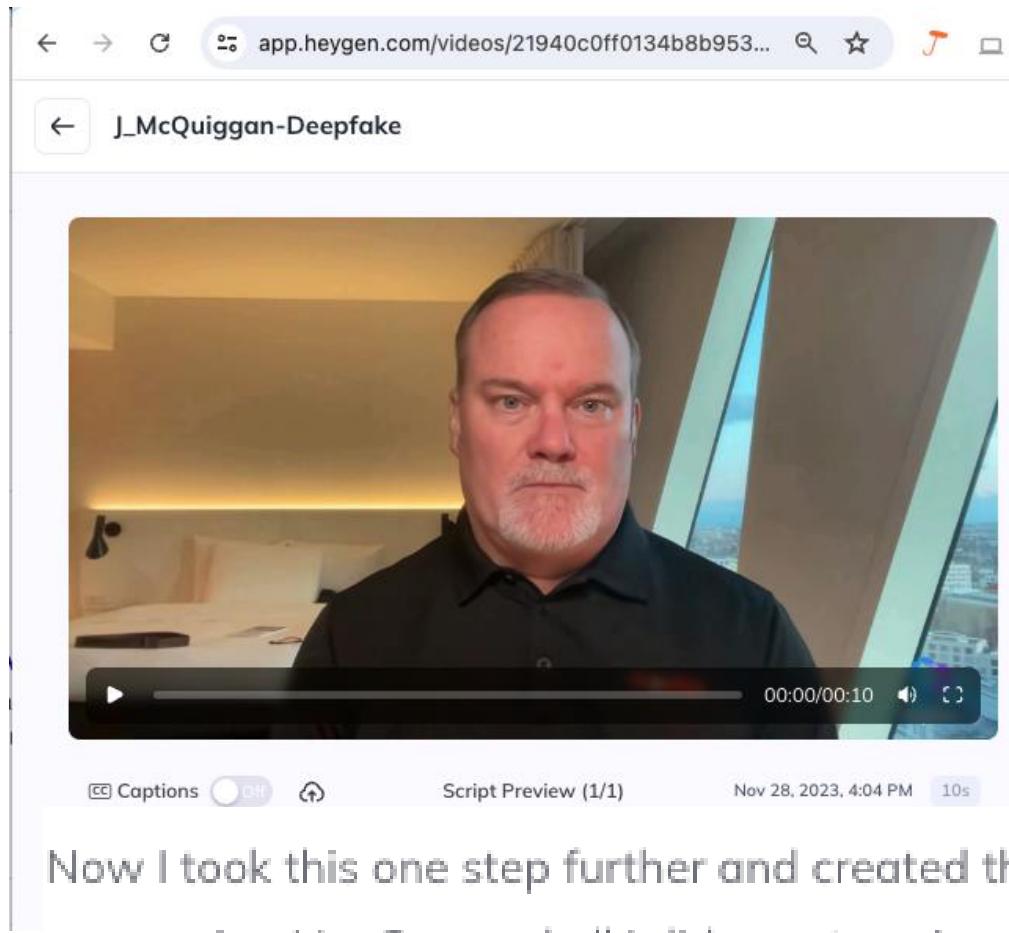
Platforms

- Smartphone apps
- SaaS
- Realtime / Stream
- Processed



DeepFaceLab - used in 95% of all deepfake videos

Modality – Text to Video (HeyGen Demo)



J_McQuiggan-Deepfake

00:00/00:10

Captions Script Preview (1/1) Nov 28, 2023, 4:04 PM 10s



Now I took this one step further and created this audio and video which I recorded from a hotel room using HeyGen and all I did was type in a script of what I wanted to say.

Creating The Deepfake

- YouTube channel – pulled video / audio
- Imported to 11Labs to generate Audio file
- Imported Headshot and audio from 11Labs to generate video on Hedra
- Good thing I asked Andrada for permission!



Dark Web Activity

14 June

© 239 edited 12:35

Black Market Ⓜ

Black Market Ⓜ Plus Plan

🔥 New Released

~~DeepFake AI~~

- The most advanced deep fake video impersonation application using the latest DeepFake AI technology.
- Supported on Windows machine with GPU and minimum 8GB RAM.
- Simply upload any person photo and let the DeepFake AI make it live with enhanced 3D dimensions following your text scripts expressions, movements and voice for the high resolution video generation.
- Best for generating your own fake / clone video statement and conference telling about anything based on your text scripts with your own preferred voice cloning module.
- The new era of video spoofing, love scamming and false statement spreading.
- Unlimited high resolution deep fake video generations.

Bundle Package Fee:

Lifetime = 🇺🇸 USD160 / 💯 USDT160

Black Market ® Premium Plan

Hot Selling

- ✓ The most advanced deep fake video impersonation tool using well known DeepFake AI model.
- ✓ Simply upload any person photo and let the DeepFake AI make it live following your expressions, movements and voice for the high resolution video generation.
- ✓ Best for generating your own fake video statement and conference telling about anything that you want.
- ✓ The new era of video spoofing, love scamming and false statement spreading.
- ✓ Unlimited high resolution deep fake video generation.

Subscription Fee:

1 month = USD60 / USDT60

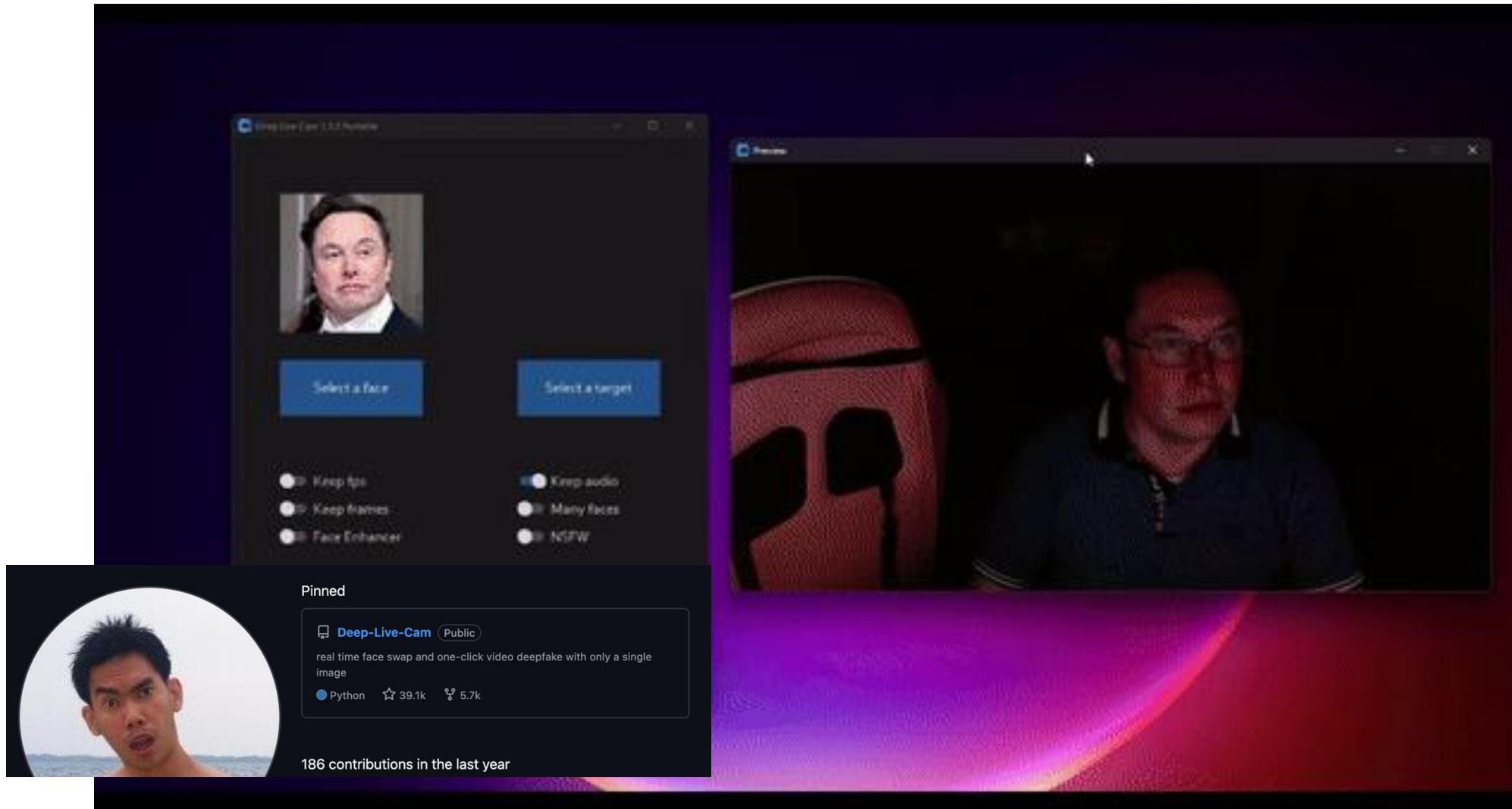
3 months = USD150 / USDT150

6 months = USD250 / USDT250

Lifetime = USD400 / USDT400

Source: <https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/surging-hype-an-update-on-the-rising-abuse-of-genai>

Deepfake & Webcams - LIVE





Synthetic Identities

Synthetic Identities – Sock Puppet Accounts

Random Face Generator (This Person Does Not Exist)

Generate random human face in 1 click and download it! AI generated fake person photos: man, woman or child.

Gender: Any Age: Any Ethnicity: Any Refresh Image

this-person-does-not-exist.com



<https://this-person-does-not-exist.com>



Fake Person Generator To protect your real information from being leaked

Related Links

- Gmail Generator
- Random Address
- Random Phone Number
- Postcode Finder
- BIN Generator
- Employment Info Generator
- Identity Generator
- User Profile Generator
- IMEI Generator
- User Face Generator
- Nickname Finder
- Temporary Mail
- Gamertag Generator

Custom Generate

Gender: Random Age: Random State: Random City: Random Generate

Gender: male
Race: Black
Birthday: 5/8/1989 (35 years old)
Street: 3161 Whitetail Lane
City, State, Zip: Dallas, Texas(TX), 75244
Telephone: 469-296-7008
Mobile: 817-675-9384



Lionel R Sanderson

<https://www.fakepersongenerator.com>

Synthetic Identity

Security Firm Discovers Remote Worker Is Really a North Korean Hacker

Security awareness company KnowBe4 noticed something was fishy when the employee's company-issued Mac began loading malware.

Jul 23, 2024



- Fastest growing type of fraud
- >\$1b in 2023
- Expected \$40b by 2027
- Face swaps Injection attacks increased 704% in H2 vs H1 2023
- Yahoo Boys = Romance Scams
- NK hired employees in Tech orgs
- Now that's an interesting story...

KnowBe4 & Synthetic Identities

X Post Button

23

How a North Korean Fake IT Worker Tried to Infiltrate Us

Jul

Stu Sjouwerman

X Post

Share

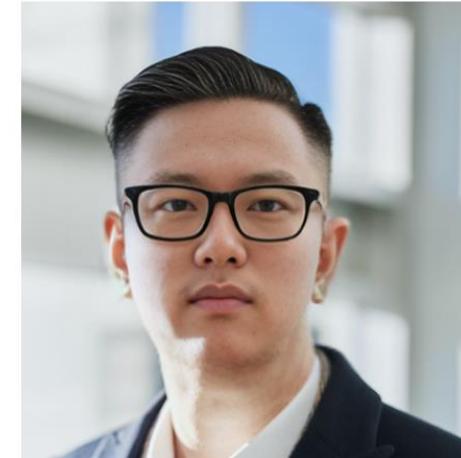
Share 443

Incident Report Summary: Insider Threat

First of all: No illegal access was gained, and no data was lost, compromised, or exfiltrated on any KnowBe4 systems. This is not a data breach notification, there was none. See it as an organizational learning moment I am sharing with you. If it can happen to us, it can happen to almost anyone. Don't let it happen to you.

We wrote an FAQ, answering questions from customers. Story updated

7/27/2024.

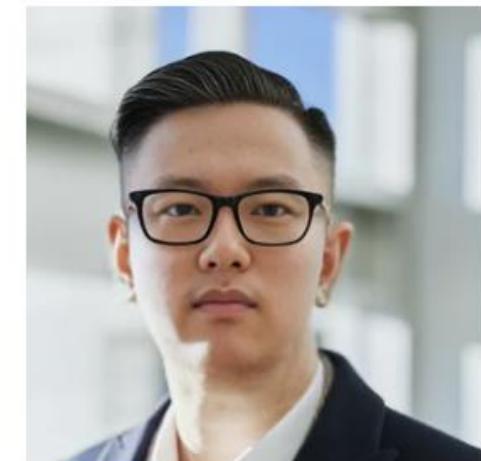


TLDR: KnowBe4 needed a software engineer for our internal IT AI team. We posted the job, received resumes, conducted interviews, performed background checks, verified references, and hired the person. We sent them their Mac workstation, and the moment it was received, it immediately started to load malware.

KnowBe4 & The North Korean Employee

Stock Photo

- Applied for Principal Software Engineer with a fake identity, appeared as a US citizen of Asian descent.
- Claimed education in Hong Kong and work experience at well-known US companies.
- Passed interviews, technical, reference, and background checks, then hired.
- Received Apple laptop with FIDO-enabled Yubikey, requested laptop to be shipped to a new location.
- July 15, 2024: Attempted to install password-stealing malware but failed multiple times.
- Generated EDR alerts; KnowBe4 SOC contacted employee who made excuses and refused an audio session.
- Laptop isolated 25 minutes after first alert.
- Incident data shared with FBI, confirmed as a North Korean fake employee.
- Public blog post released; story went viral with extensive press coverage.
- Multiple companies shared similar experiences, leading to webinars, a whitepaper, and educational articles.



AI-Photo
Blending of NK emp.



10 Things KnowBe4 Learned from the NK Hiring Event



1. Trained recruiters to spot DPRK resume red flags.
2. Provided recruiters phone carrier lookup tools.
3. Phone-based screening required and not just email
4. Recruiters search for applicant's public social media profiles.
5. Updating identity verification to government standard.
6. Video interviews require camera on without background filters.
7. Recruiters trained to ask casual, location-based questions.
8. CISO consults if suspicion arises during interviews.
9. Equipment shipped only to verified addresses or UPS stores.
10. Internet searches of addresses on suspicious resumes.

Source: <https://blog.knowbe4.com/north-korean-it-worker-threat-10-critical-updates-to-your-hiring-process>



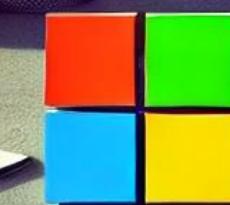
An Algo-rithim



Edge
on edge?
on the way?

Never
relaxed?

Microsoft



Microsoft



Detection / Prevention



Deepfake Scams

Fraud Loss Projections

Expected
~\$40b by 2027

Increase in Deepfake Cases

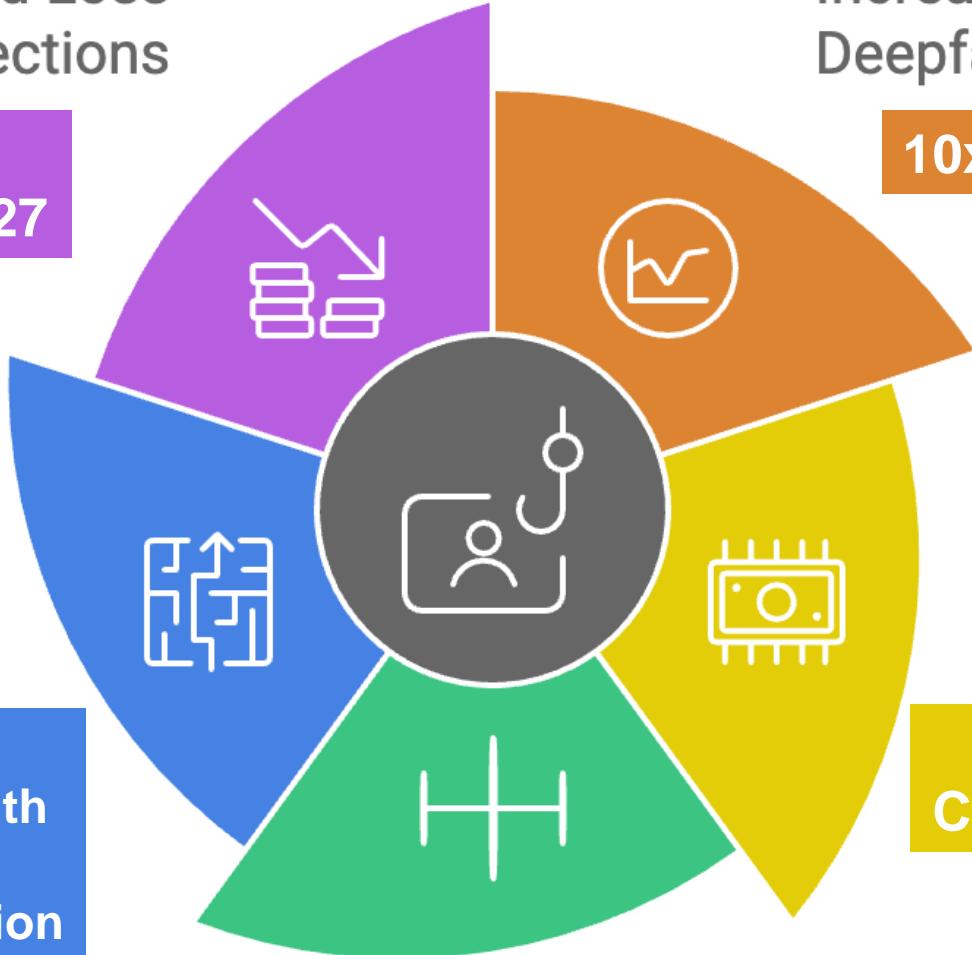
10x Increase

Business Awareness

25% execs unfamiliar with
32% doubt users detection

Targeted Industries

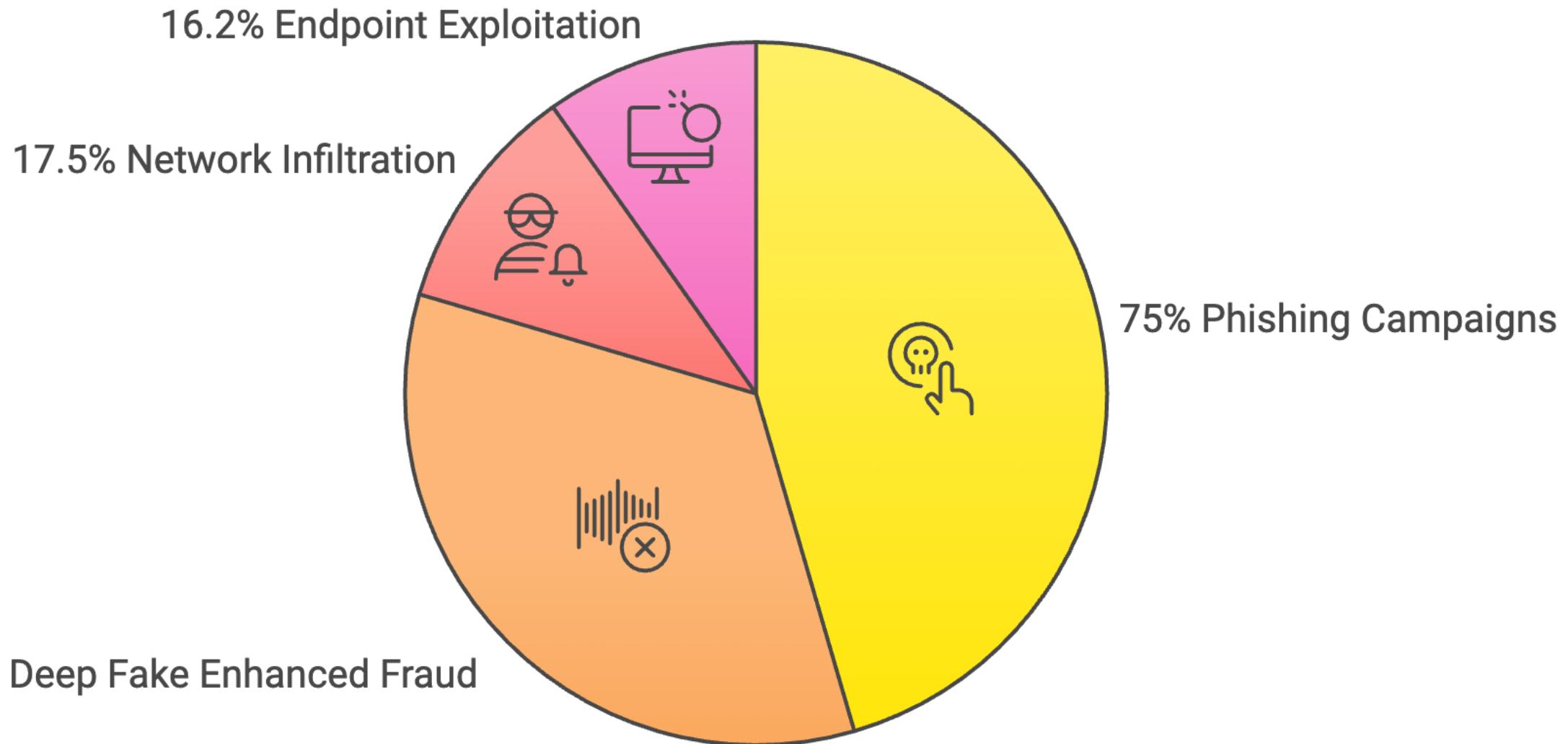
88% in
Crypto Sector



Regional Impact

1740% Increase in NAM

CISO Survey – AI Threats – Team 8



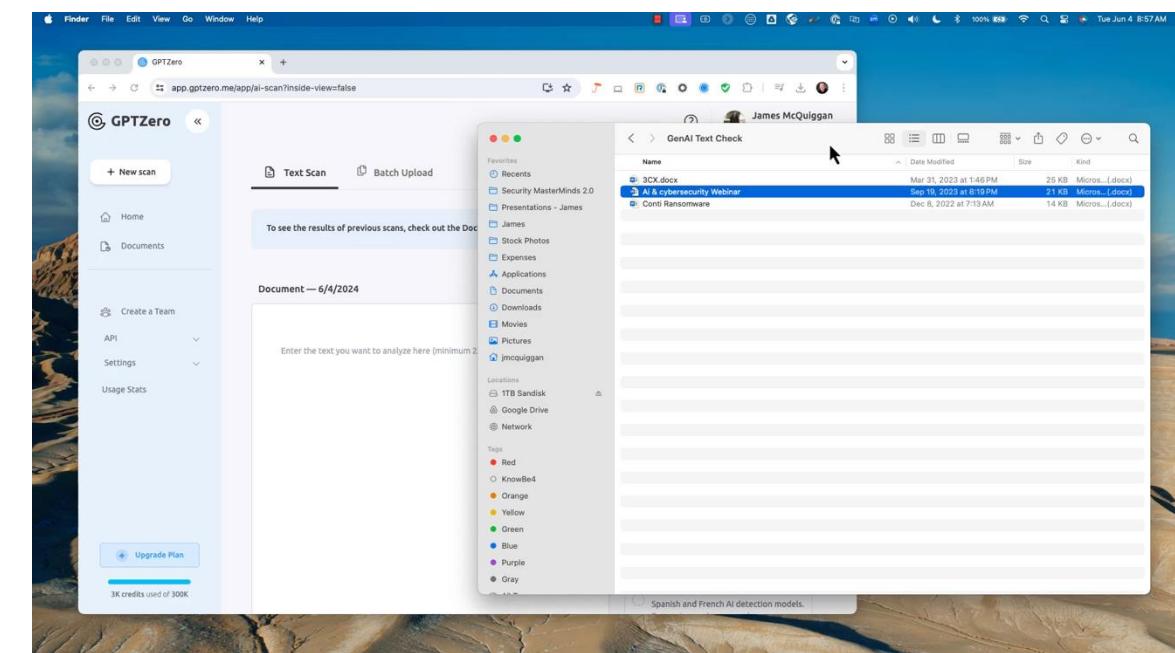
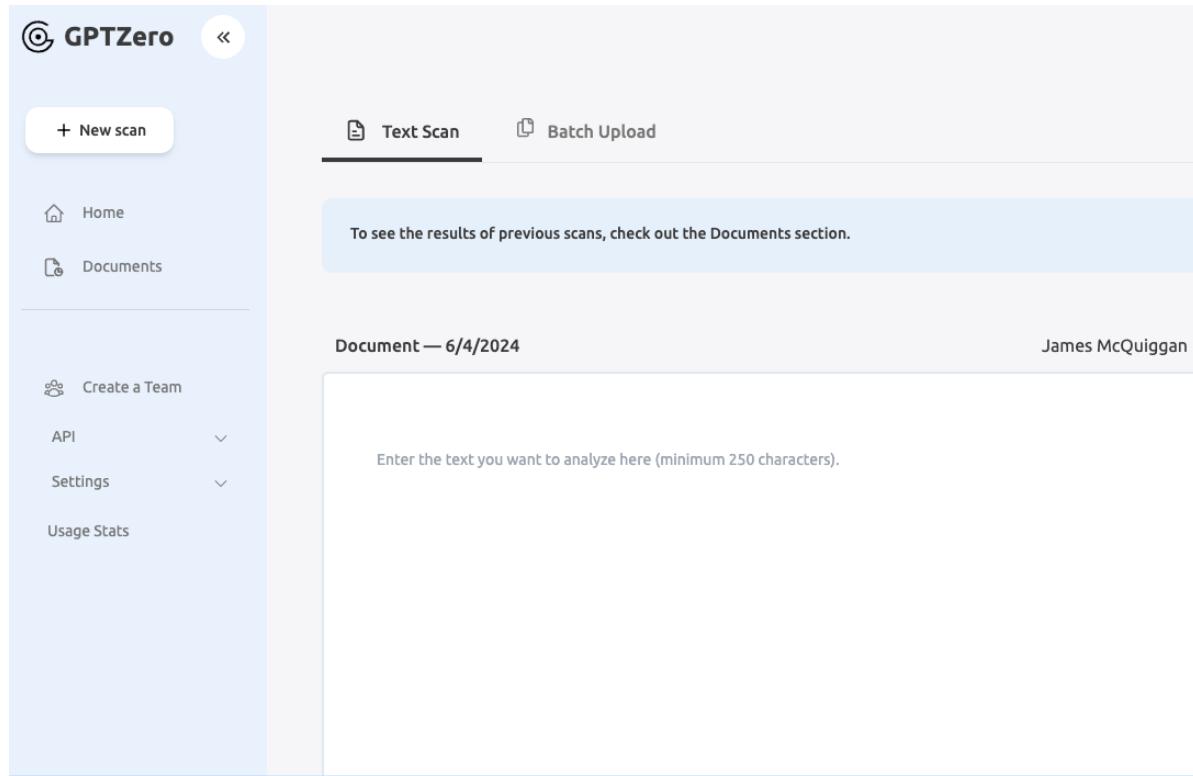
Basic Detection

- The criminals have leveled up
- We need to continue the same path...
with some adjustments
- People, Processes, Technology
- People: If interacting, consider code words or ask unusual questions
- Technology: Detectors
- Processes: Frameworks



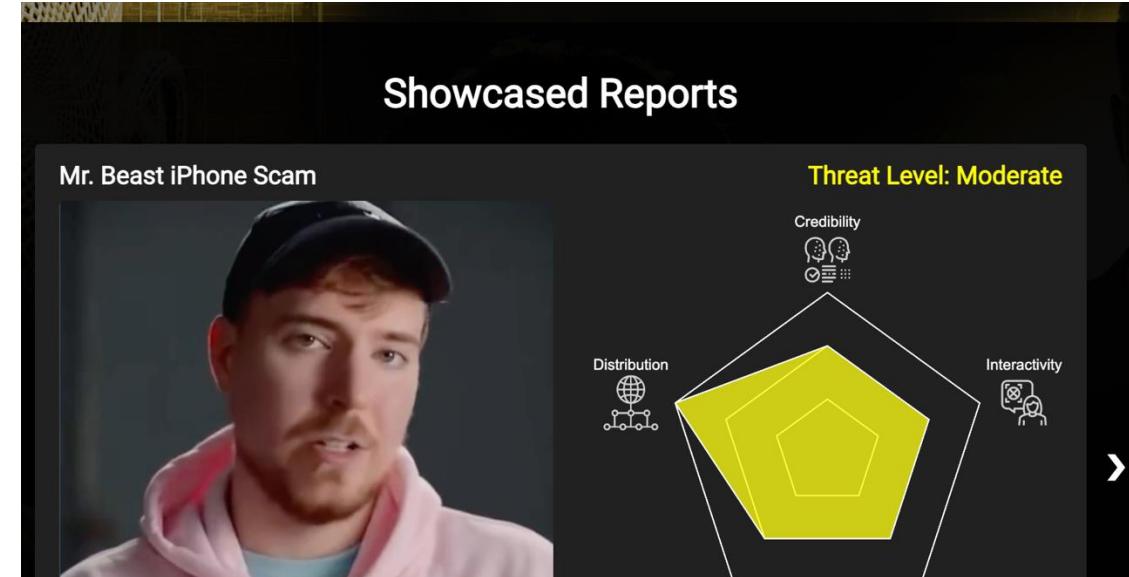
GPTZero – GenAI Synthetic Text Detection

- <https://app.gptzero.me/app/ai-scan>



Synthetic Video Detection Challenges

- Non-real time
- Not full-proof
- No standard detection method yet
- Generation tech advances outpace detection tech
- False Positives are plentiful
- Still requires manual labor



ST Engineering launches Einstein.AI deepfake detector

By Adam Campbell Last updated September 5, 2024





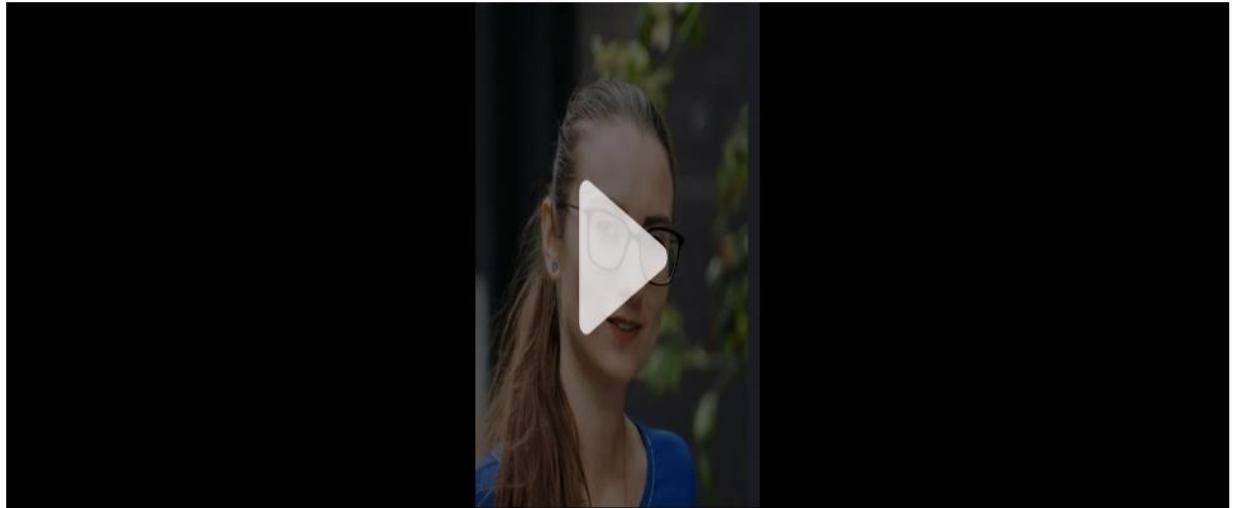
NO DEEPFAKE DETECTED



Name: af66f232-b4ac-4a35-a2dd-d6c93655ba68.mp4

Size: 9.5 MB

Deepware aims to give an opinion about the scanned video and is not responsible for the result. As Deepware Scanner is still in beta, the results should not be treated as an absolute truth or evidence.



Model Results

Avatarify: NO DEEPFAKE DETECTED(43%)Deepware: NO DEEPFAKE DETECTED(0%)Seferbekov: NO DEEPFAKE DETECTED(1%)Ensemble: NO DEEPFAKE DETECTED(0%)

Video

Duration: 37 sec**Resolution:** 512 x 512**Frame Rate:** 30 fps**Codec:** h264

Audio

Duration: 37 sec**Channel:** mono**Sample Rate:** 48 khz**Codec:** aac

James HeyGen Video - Deepware

Deepware aims to give an opinion about the scanned video and is not responsible for the result. As Deepware Scanner is still in beta, the results should not be treated as an absolute truth or evidence.

 **DEEFAKE DETECTED**

 Name: James -BSidesCPH.mp4 | User: 202
Size: 5.5 MB | Source:



Model Results

Avatarify: DEEFAKE DETECTED(94%)

Deepware: NO DEEFAKE DETECTED(0%)

Seferbekov: NO DEEFAKE DETECTED(31%)

Ensemble: NO DEEFAKE DETECTED(4%)

Video

Duration: 9 sec

Resolution: 1920 x 1080

Frame Rate: 25 fps

Codec: h264

Audio

Duration: 9 sec

Channel: stereo

Sample Rate: 48 khz

Codec: aac

O:	
n:	9 sec
I:	stereo
Rate:	48 khz
	aac

Andrada's Audio File

 detect.resemble.ai/results/0b7e6bac1708987c39e00b3d2805fd0c

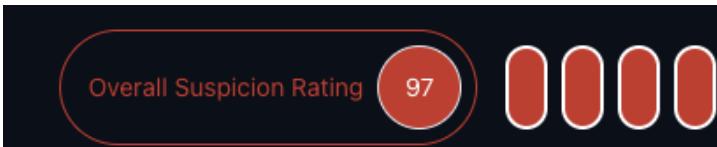
Resemble Detect

Detect deepfake audio from any source with our powerful AI Model. [Try it out yourself →](#)



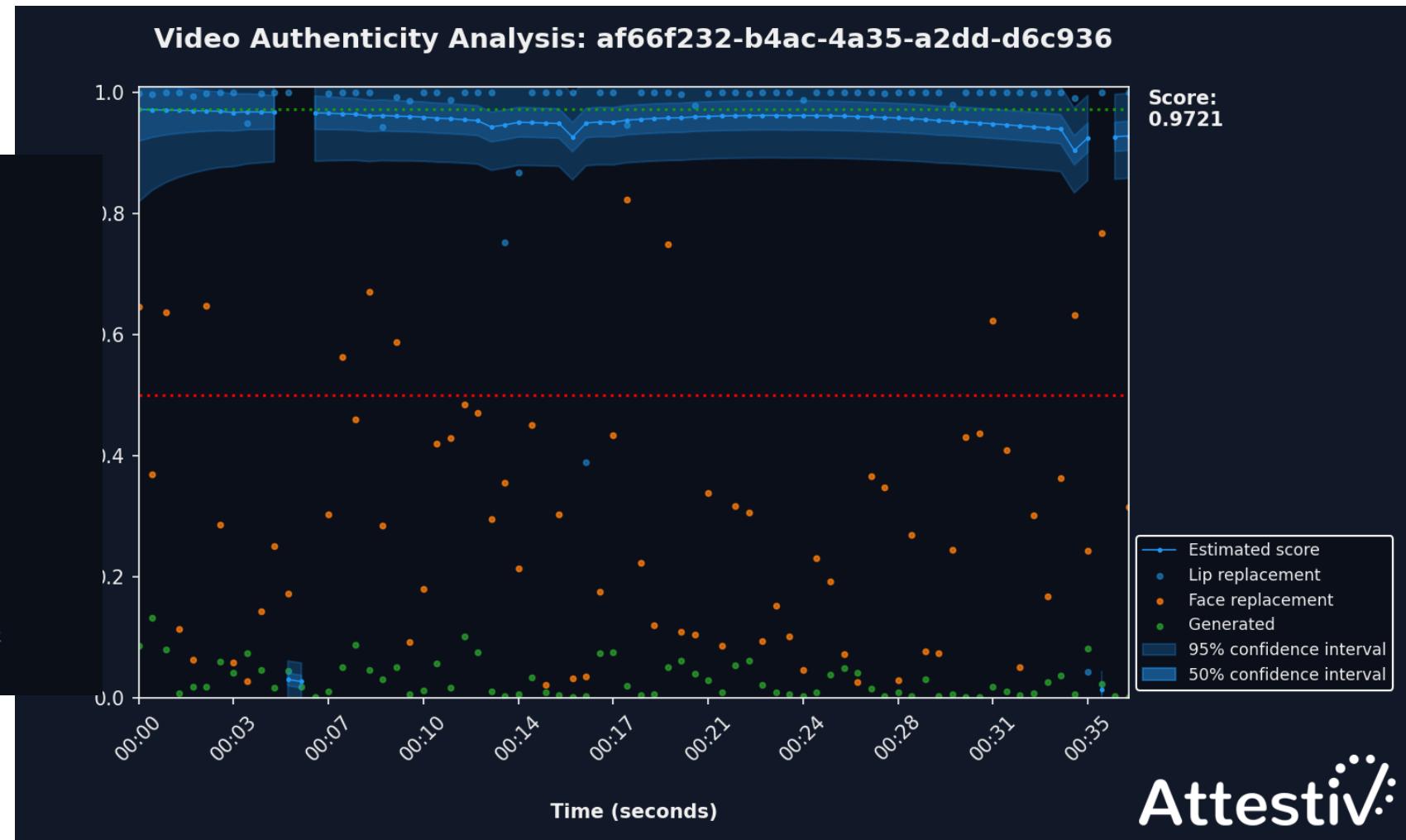
Result: Fake

<https://detect.resemble.ai/results/0b7e6bac1708987c39e00b3d2805fd0c>



Our analysis of the video shows that the overall suspicion rating is 97, which means that parts of the video are **VERY SUSPICIOUS***

* Please note that the Overall Suspicion Rating (OSR) is an estimate based upon our analysis, and not a guaranteed or definitive statement of the presence of editing or deepfakes. [Please click here](#) to learn more about the OSR and Attestiv's stance on responsible AI. You can also learn more about our analysis and interpreting our results by visiting our [FAQ page here](#).



10 Best AI DeepFake Detector Tools



Additional Deepfake Detectors

- Sentinel.ai – Requires a demo
- Sensity.ai – Requires Corporate email and detailed reason to use their platform
- Oz Forensics – Facial recognition
- DuckDuckGoose – demo required
- Deepware – free tool – False Positive
- Attestiv.com – free tool – 65%



Process: 3 Questions for Email / Synthetics



Process: Synthetic Video Tips (VeSSPER)

Verify

- Ask questions or get them to do something unpredictable like writing a specific word on paper and showing it on camera.

Skepticism

- Be cautious if someone you've only met online requests money, personal information or any other sensitive details.

Secure

- Use secure, encrypted apps for texting and voice

Privacy

- Protect personal information available publicly

Education

- Keep up to date with newsletters, podcasts etc.

Report

- Report it to the relevant authorities like ic3.gov or police

"Don't ask, 'Is this real?' Ask, 'Why does this exist?'" - FAIK



FAKE
OR
REAL?

What Should We Be Asking?



Is this a deepfake?



Consider these questions...

What Should We Be Asking?



Why does this exist?



What story is it telling?



Who Benefits?



What are the possible goals?

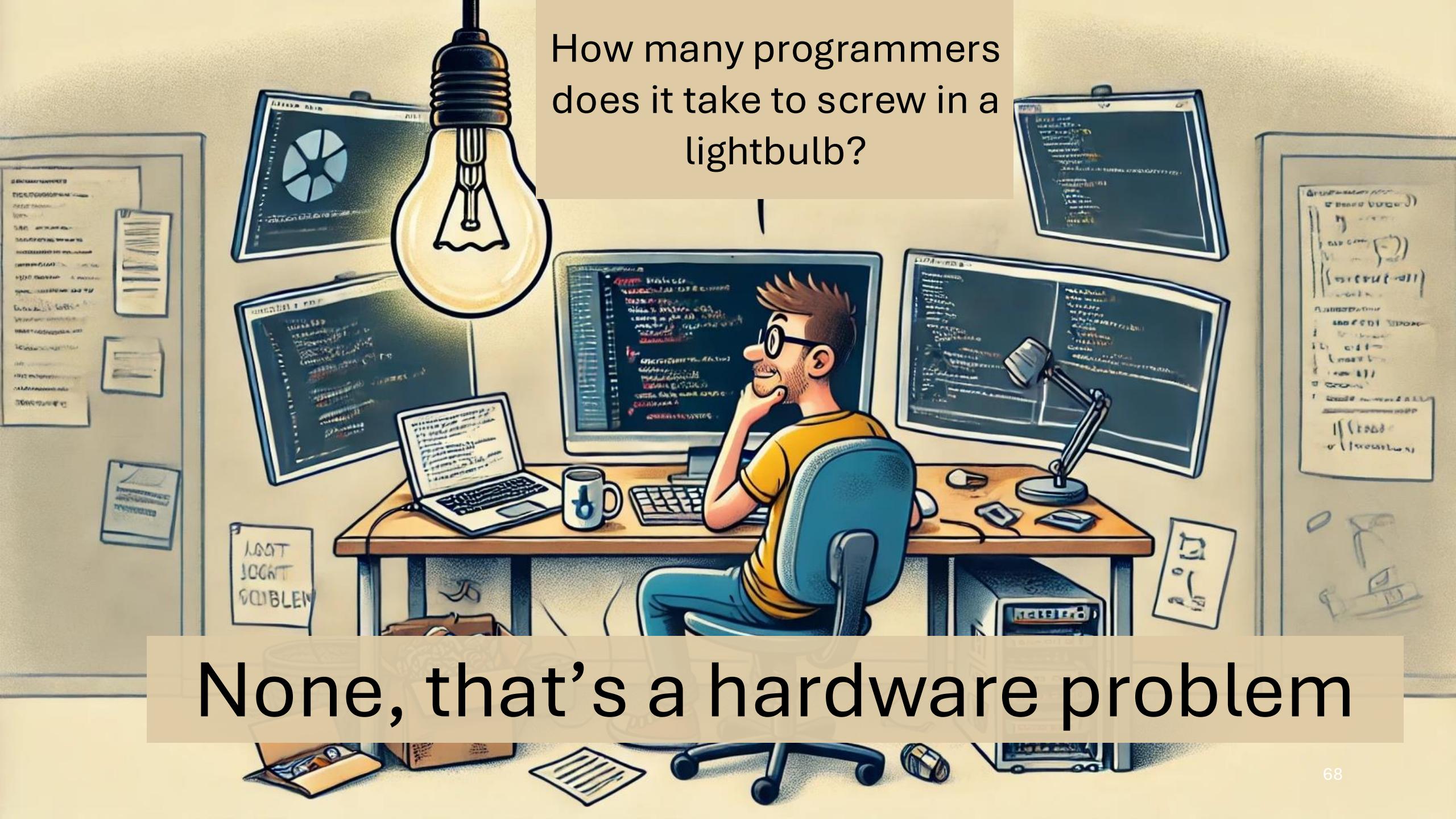
Apply the FAIK Factor Framework

F Freeze & Feel

A Analyze the Narrative & Emotional Triggers

I Investigate (claims, sources, etc.)

K Know, confirm, and keep vigilant



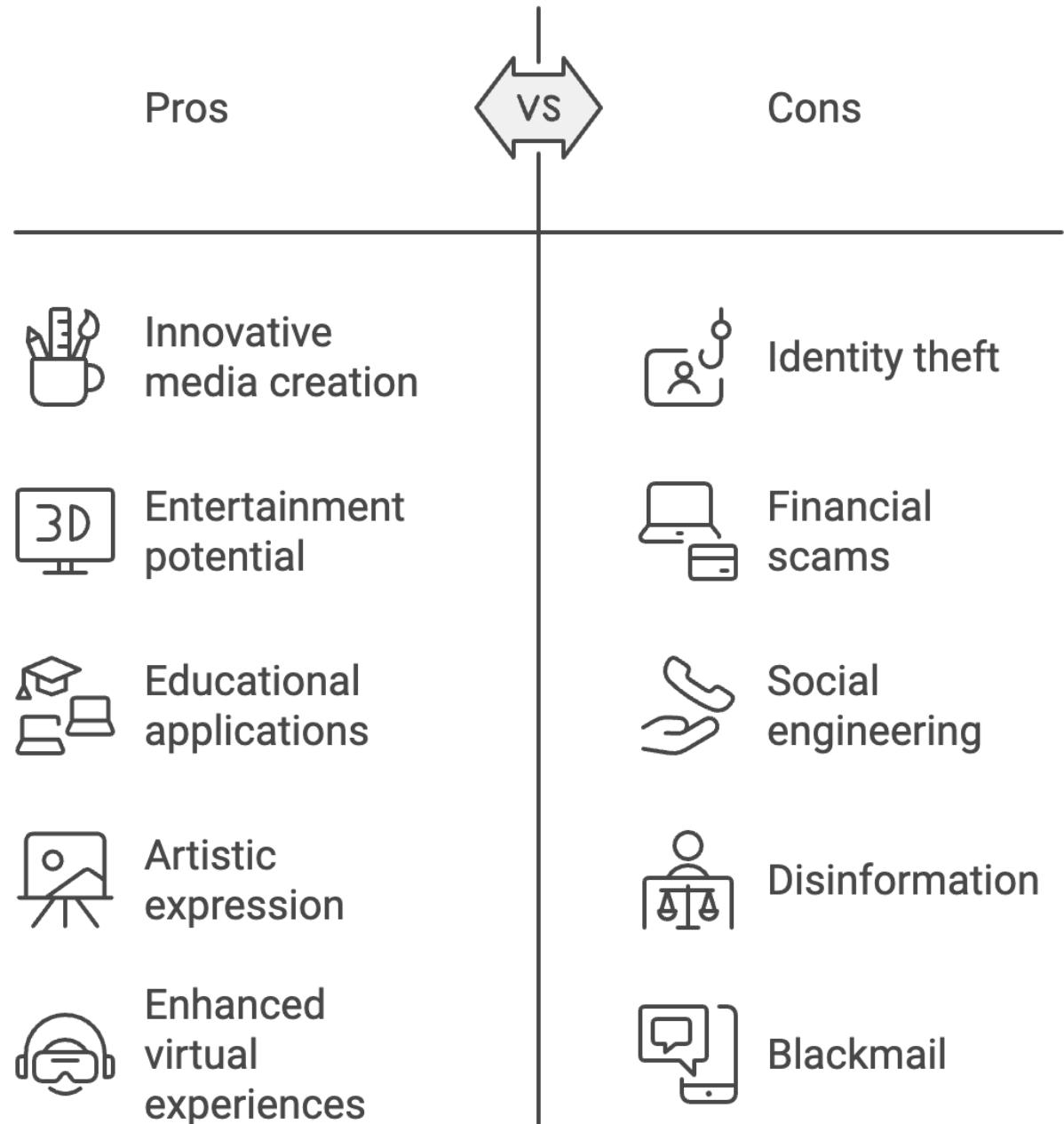
How many programmers
does it take to screw in a
lightbulb?

None, that's a hardware problem

Final Thoughts / Wrap-up



Deepfakes Duality



Takeaways

AI is an incredible
tool available to all –
Ensure we're
educating everyone.
Politely Paranoid

Be aware of AI
Hallucinations,
Biases and
Deepfakes
Trust AND Verify

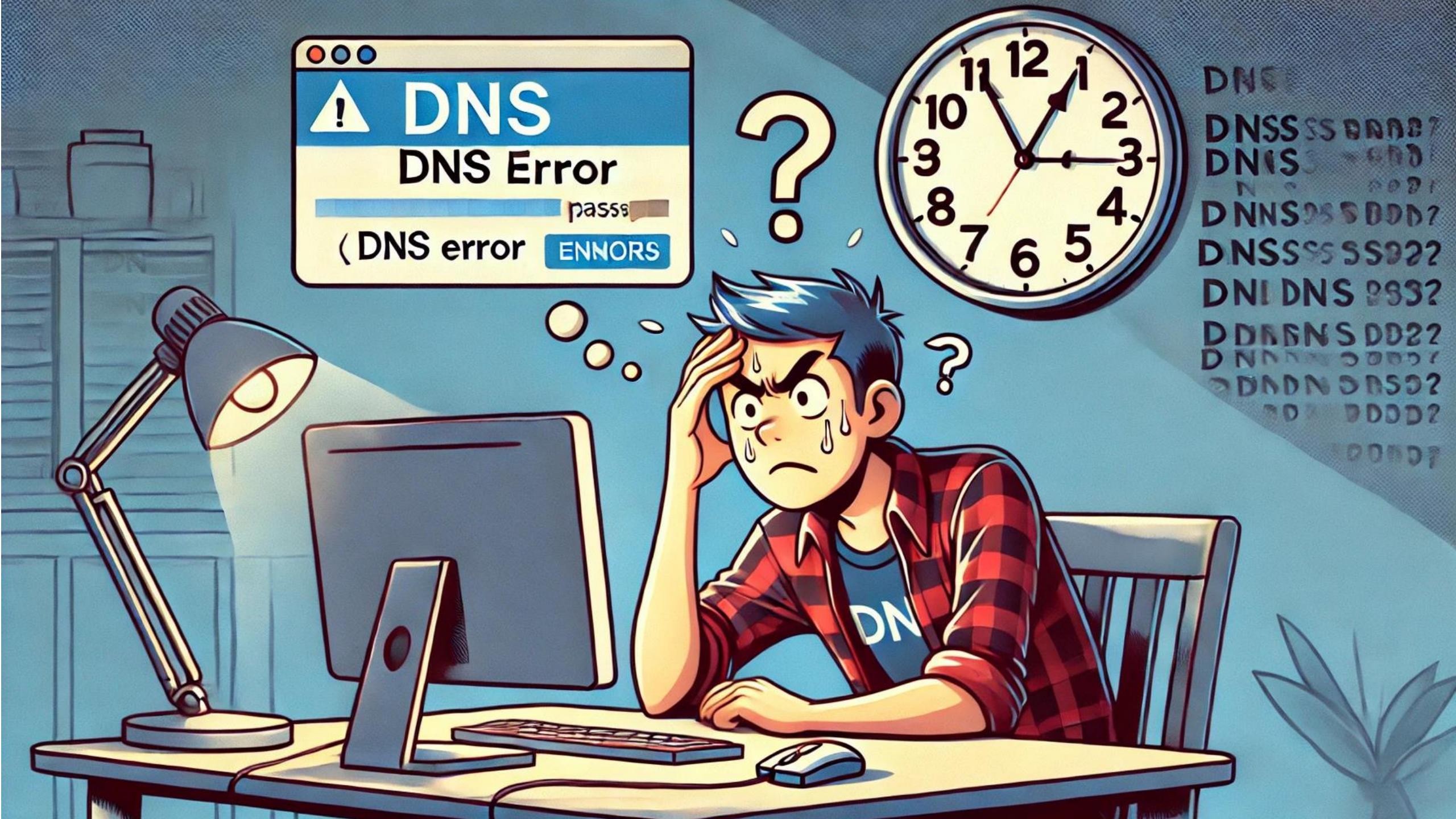
The Phishing game
hasn't changed.
Be aware, don't
rush, check links,
be skeptical

Final Thoughts

**TRUST
BUT
VERIFY**

Politely
Paranoid

Healthy
Skepticism



DN
DNSSSS BBBB?
DNIS
NN S DDD?
DNNSSSS BBBB?
DNSSSSSSSSSSSS?
DNI DNS BBB?
DDBBNS DD22
DNNNN SSSSS?
DNDN DR5D2
DD DDDDD?

Thank You For Your Attention

James R. McQuiggan, CISSP, SACP

Email: jmcquiggan@knowbe4.com

KnowBe4 Blog: blog.knowbe4.com



Connect with Me!



LinkedIn [jmcquiggan](https://www.linkedin.com/in/jmcquiggan)



X [@james_mcquiggan](https://twitter.com/james_mcquiggan)



Website jamesmcquiggan.com

☰ YouTube

Search

James McQuiggan, CISSP, SACP
35 subscribers

HOME VIDEOS PLAYLISTS CHANNELS ABOUT

Dad Jokes & Cyber Stories ► PLAY ALL

A New Business 0:37 James McQuiggan, CISSP, SACP 6 views • 4 days ago

Cybercrime 0:41 James McQuiggan, CISSP, SACP 23 views • 12 days ago

Most Secure Woman 1:01 James McQuiggan, CISSP, SACP 21 views • 2 weeks ago

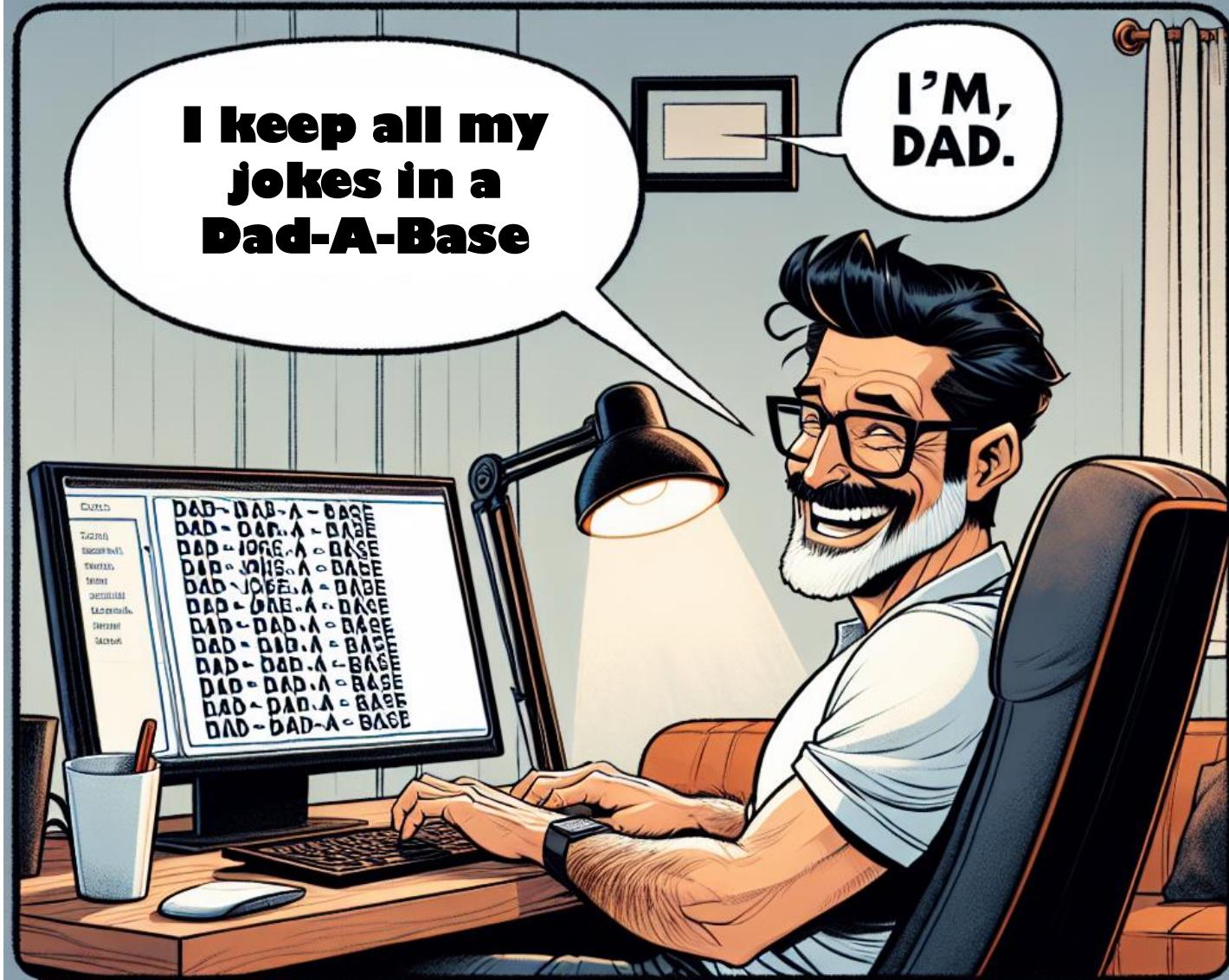
Sick Webpages 0:22 James McQuiggan, CISSP, SACP 13 views • 1 month ago

Library History Your videos Watch later Liked videos Dad Jokes & Cyber St...

YouTube: James McQuiggan and Dad Jokes

<https://www.youtube.com/@JamesMcQuigganCISSP>

Yes... I have a
way of keeping
track of my
Dad Jokes



Resources

- Resemble Audio - <https://detect.resemble.ai/>
- Deepware Detector - <https://scanner.deepware.ai/>
- HeyGen Video - <https://app.heygen.com>
- Hedra - <https://www.hedra.com>
- 11Labs – <https://elevenlabs.io/>
- PlayHT - <https://play.ht/>
- **Presentation:** <https://github.com/jmcquiggan/presentations>



THANK YOU!