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. 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presentation, and the following written materials, contain KnowBe4's proprietary and confidential information and is not to be published, duplicated, or distributed to any third party without KnowBe4's prior written consent.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 Information in this presentation speaks only as of the date that it was prepared and may become incomplete or out of date; KnowBe4 makes no commitment to update such information.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.

Product and Tool References Disclaimer: Any products, tools, software, or services referenced, demonstrated, or mentioned in this presentation are provided for informational and educational purposes only. The inclusion of such references does not constitute an endorsement, recommendation, or approval by KnowBe4 or the presenter. KnowBe4 and the presenter have not received any compensation, consideration, or other financial benefit from the referenced product or service providers. Attendees should conduct their own independent evaluation and due diligence before using any referenced products or services. KnowBe4 and the presenter disclaim any responsibility or liability for the performance, security, or suitability of any referenced products or services.

DISCLAIMER

Is it Live or is it Memorex?
Memorex is a deepfake!

@31337magician

dylan-baklor

How long do you think it took me to make these videos?

James R. McQuiggan

Professor, CTI, Full Sail University
Host, Simply Secured Podcast

CISSP. Certified Information Systems Security Professional
ISC2 Certification

H LAYER CREDENTIALING
CERTIFIED
Security Awareness and Culture Professional (SACP)

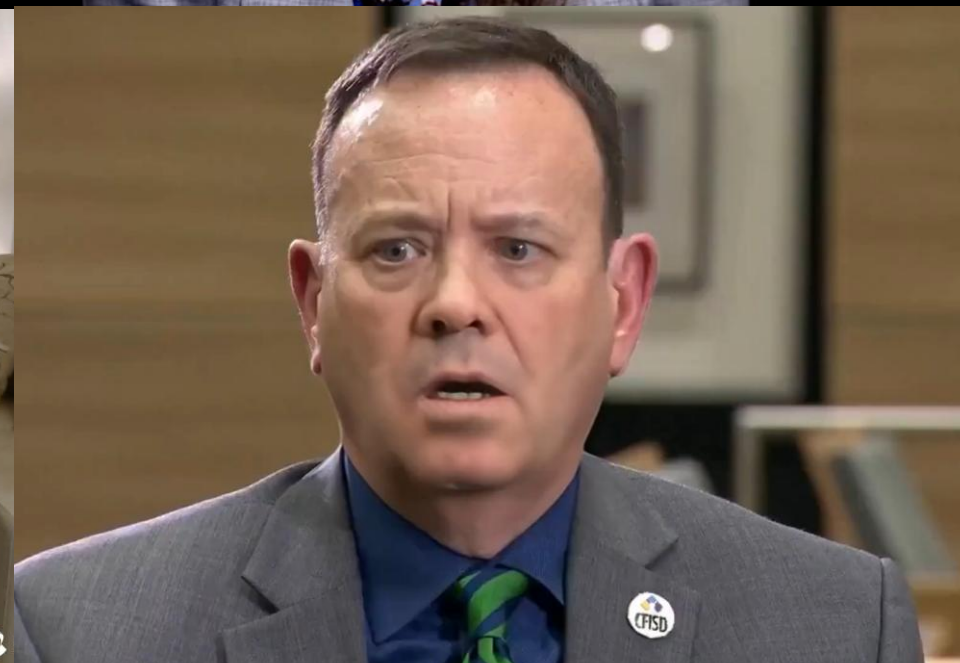
OSC
OPEN-SOURCE
CERTIFIED


SECURITY
AWARENESS

When I started messing with Deepfakes






HACKER


Why Deepfakes Matters in Cybersecurity

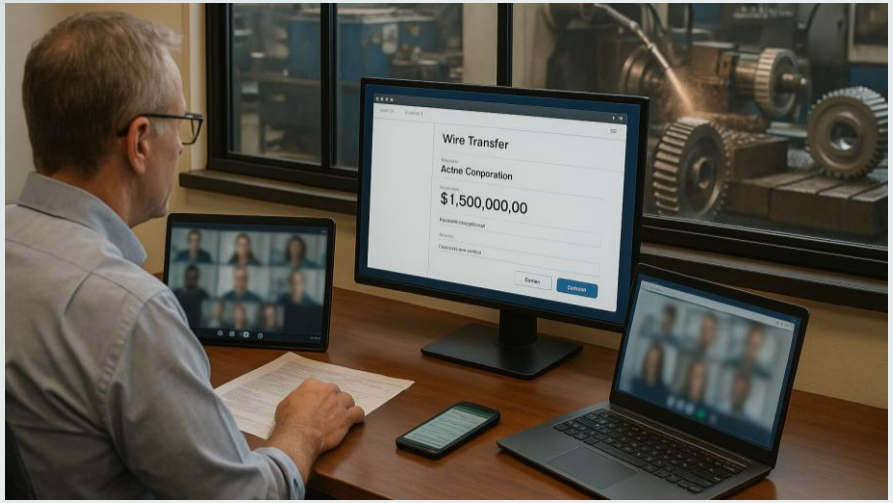
FINANCIAL TIMES

COMPANIES TECH MARKETS CLIMATE OPINION LEX WORK & CAREERS LIFE & ARTS HTSI

Cyber Security + Add to myFT

Arup lost $25mn in Hong Kong deepfake video conference scam

UK-based engineering group identified as target of fraud that used digitally cloned CFO to trick staff



Financial Fraud

Involves scams and wire transfer manipulation

Operational Disruption

Includes supply chain and IT support issues

Reputation Damage

Covers fake press releases and crisis statements

Regulatory Risk

Encompasses disclosure failures and compliance issues

A Quick Look at Synthetic Media

Which network excels?

Generator

Creates synthetic data

Discriminator

Evaluates data authenticity

Synthetic Text - Malicious LLMs

Phishing LLM – Nation state Activity - LAMEHUG

Attribution:	Attack Vector	Technical Details	Timeline
• Russian APT28 (Fancy Bear/Forest Blizzard) • ZIP archives containing three LAMEHUG variants • Hugging Face: Qwen2.5-Coder-32B-Instruct	• Phishing emails from compromised (BEC) government account • Impersonating ministry officials • Targeting executive government authorities	• Leverages Alibaba Cloud's coding-focused LLM • Available on Hugging Face and Llama platforms • Malicious command generation with python	• First reported to CERT-UA on July 10, 2025 • This is an active, current threat

Source: https://thehackernews.com/2025/07/cert-ua-discovers-lamehug-malware.html

Recent LLM Attacks

Agentic AI coding assistant helped attacker breach, extort 17 distinct organizations

Cyber Press

World's First AI-Powered Ransomware 'PromptLock' Emerges Using GPT-OSS-20B

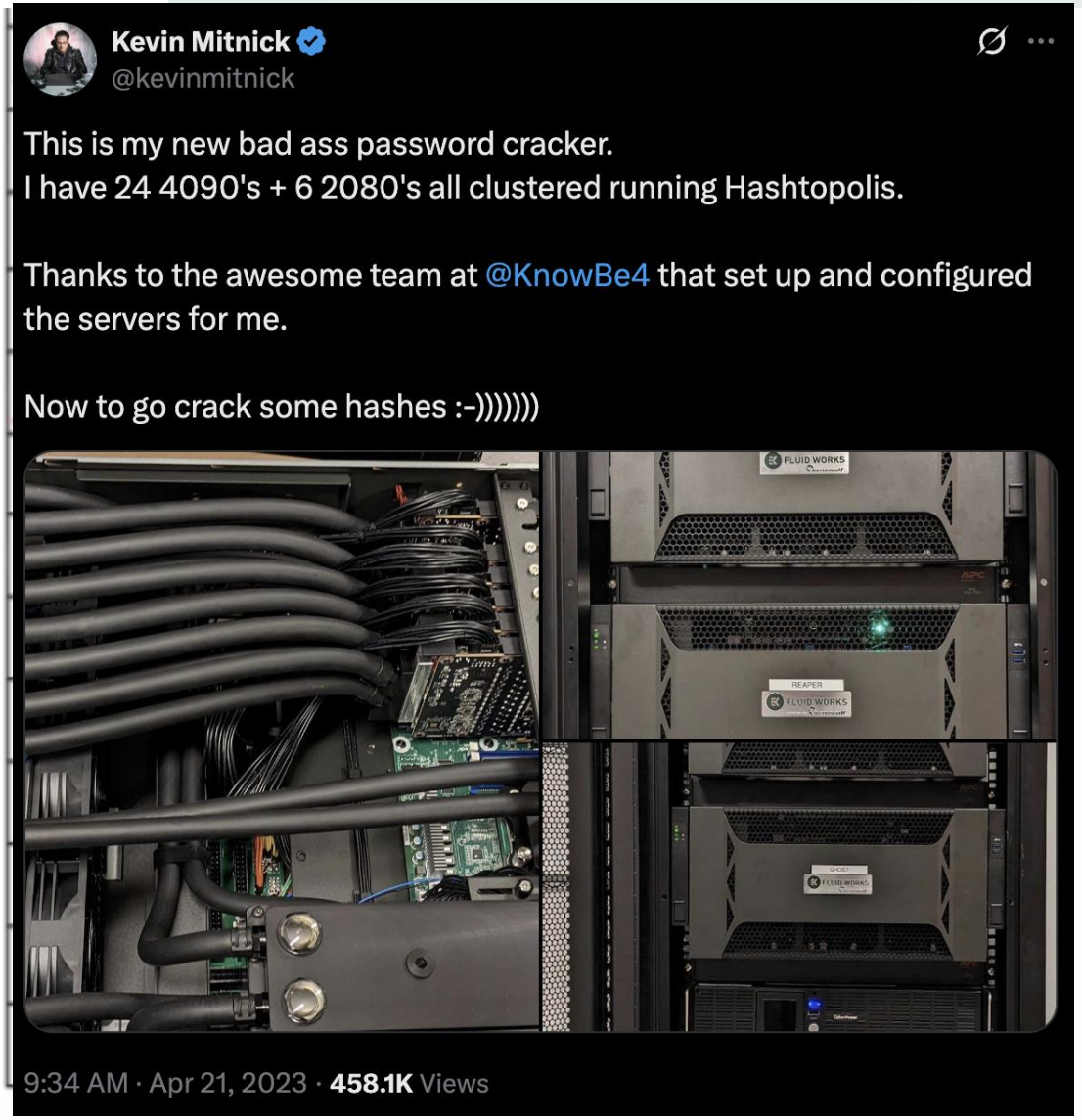
BY ANUPRIYA / AUGUST 28, 2025

Categories: Cyber Security News •• Cybersecurity •• Ransomware

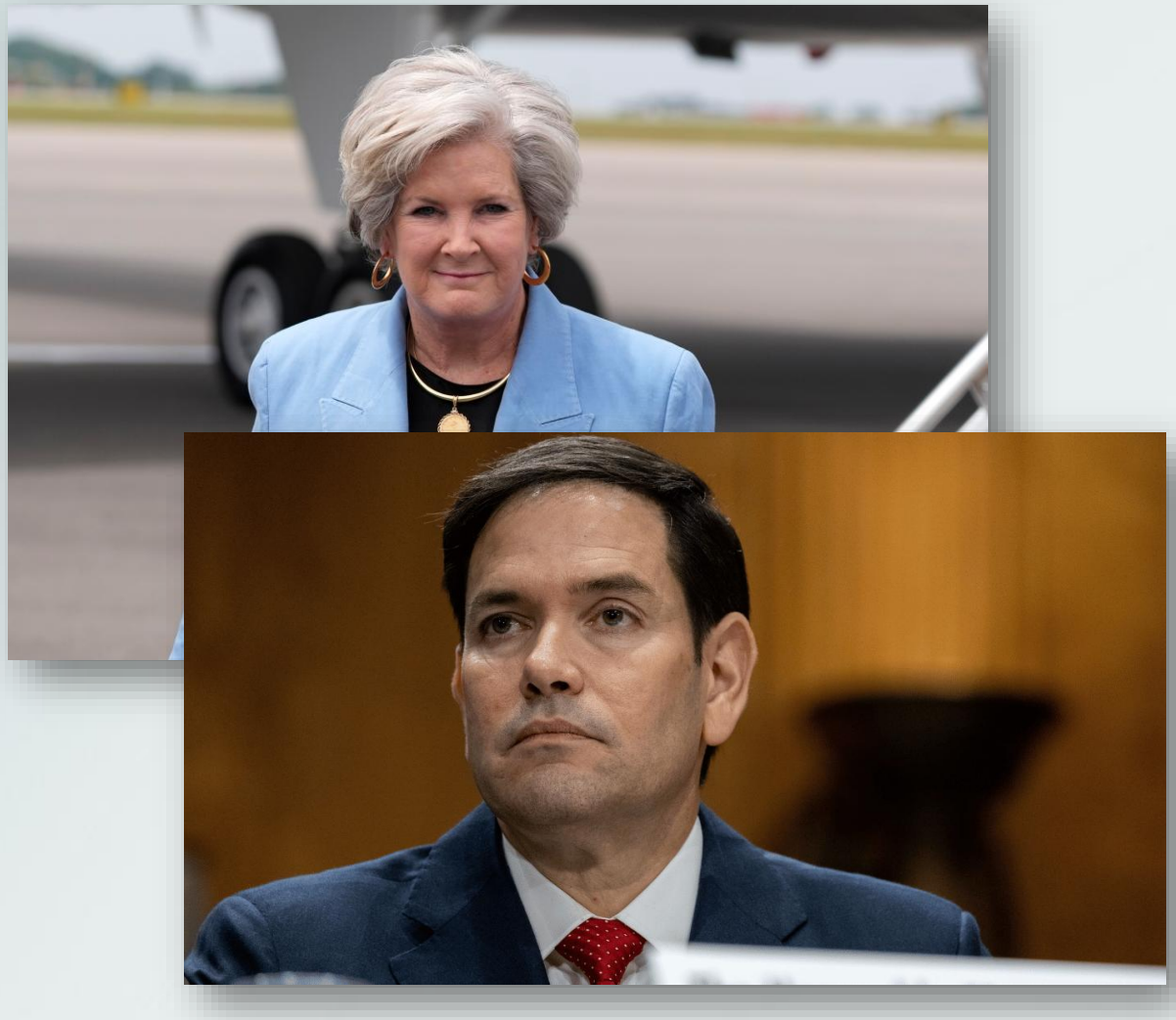
World's First AI-Powered Ransomware 'PromptLock' Emerges Using GPT-OSS-20B

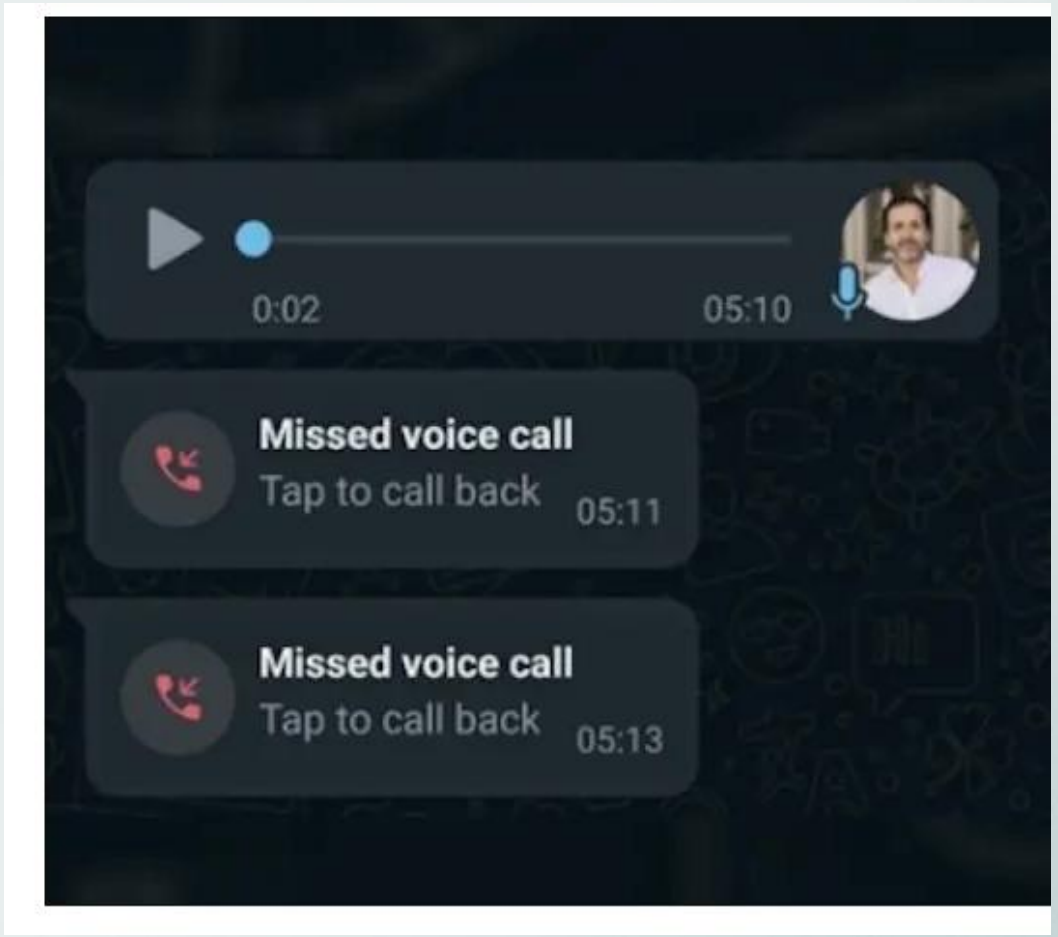
Password Cracking – Latest Results From KnowBe4

Forthehopeofitall13$
khamzatchimaev1!
Houseofdragon2894!
CastleArsenal123
LionsSuperbowl2223
Ofundamental@123A
Megan1009@@@@@@@@
#LydiaLynne2019!
Soccerrugby110$$
KasselGermany1924
RaidShadowLegendsRocks
TheDreamJob4242!
Buddycannoli888!
Myfriendsaremypower323!!



Synthetic Audio and Social Engineering





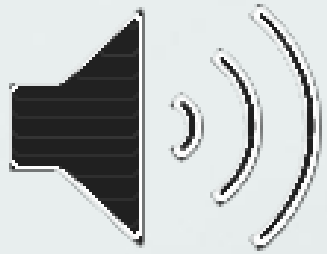
Kyle Wilhoit of Palo Alto Networks' Unit 42 division

ChatGPT + Syn. Audio + Call Center = AI Social Engineering

Using PlayHT

Dr. Gerald Auger
Simply Cyber
(and friend)

Call Center
Support
Software



Disinformation Campaigns

Hany Farid • Following
UC Berkeley Professor & GetReal Co-founder
14h • Edited •

In just the last 12 hours, we at **GetReal** have been seeing a slew of fake videos surrounding the recent conflict between Israel and Iran. We have been able to link each of these visually compelling videos to Veo 3.

It is no surprise that as generative-AI tools continue to improve in photo-realism, they are being misused to spread misinformation and sow confusion.

One simple tip-off (for now at least) is that all of these videos are either exactly eight seconds in length or composed of short (eights seconds or less) clips composited together. Why eight seconds? This is the current maximum length that Veo 3 can generate a continuous shot. Other models have slightly longer limits but 8-10 seconds is typical.

This eight-second limit obviously doesn't prove a video is fake but should be a good reason to give you pause and fact-check before you re-share.

You and 362 others 21 comments · 51 reposts

The National Guard's post

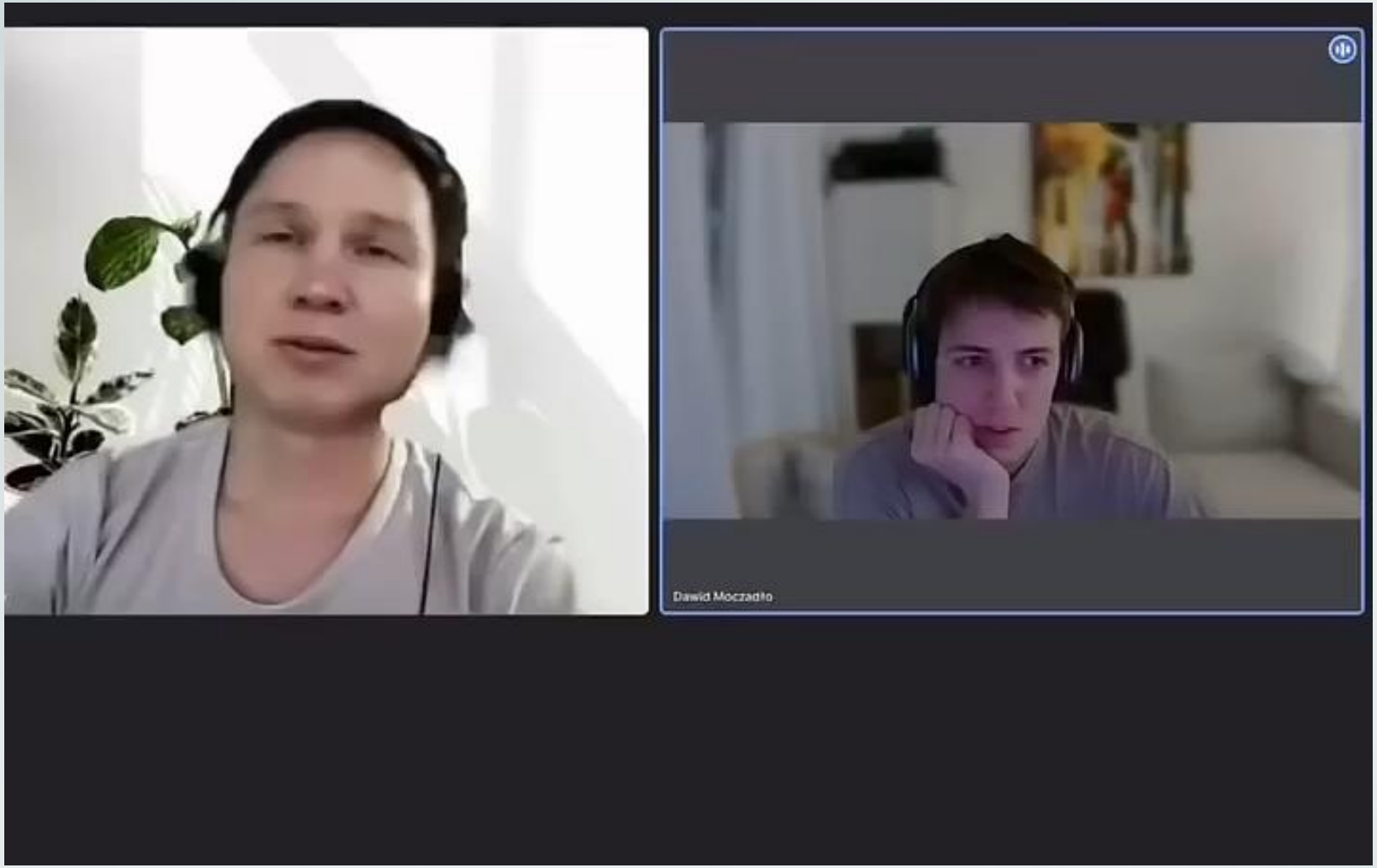
The National Guard
5d ·

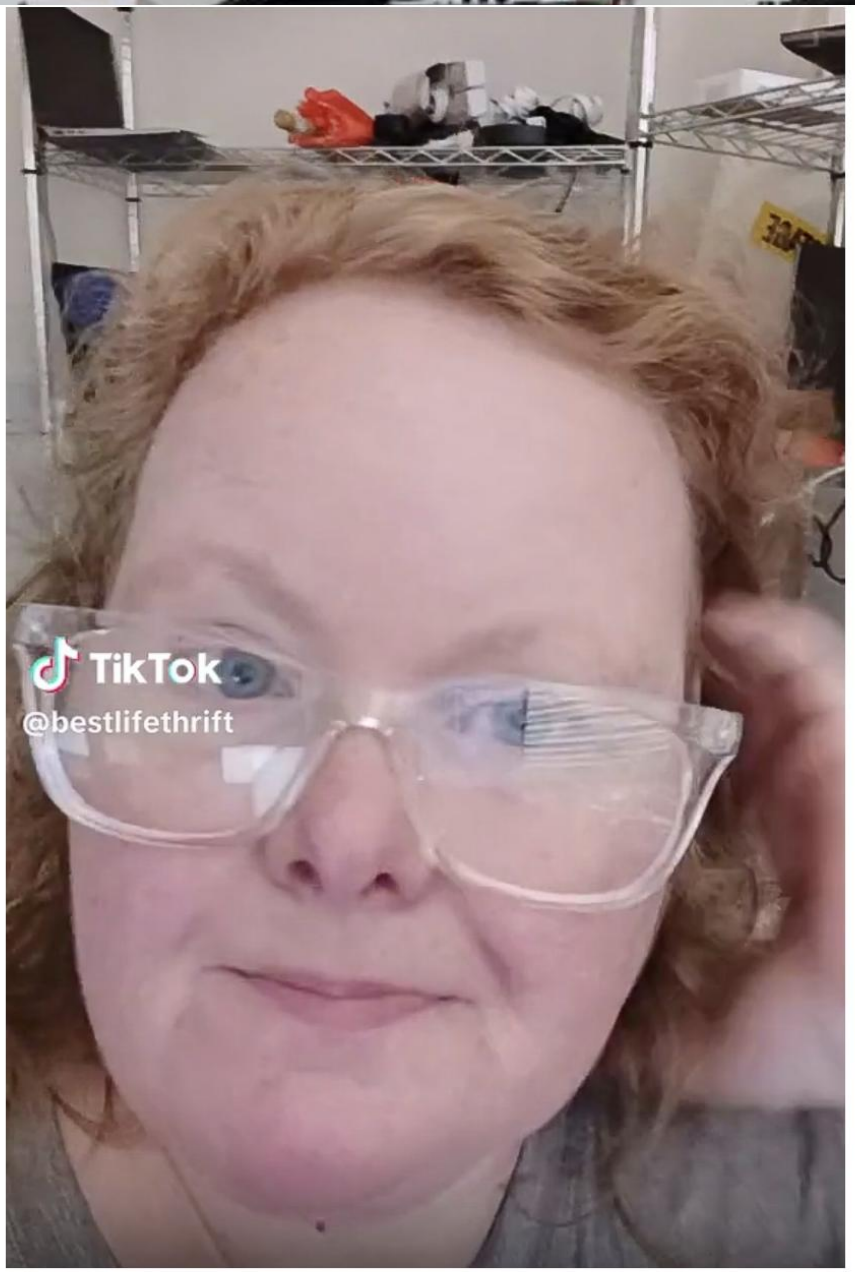
Several videos from "Bob" have been making the rounds online. They are fake. Red flags indicating a fake: The uniform name tape reads "Bob," his rank reads "E-6," and there is gibberish where "U.S. Army" should appear. In one video, the AI-generated "man" even eats a burrito through a mask.

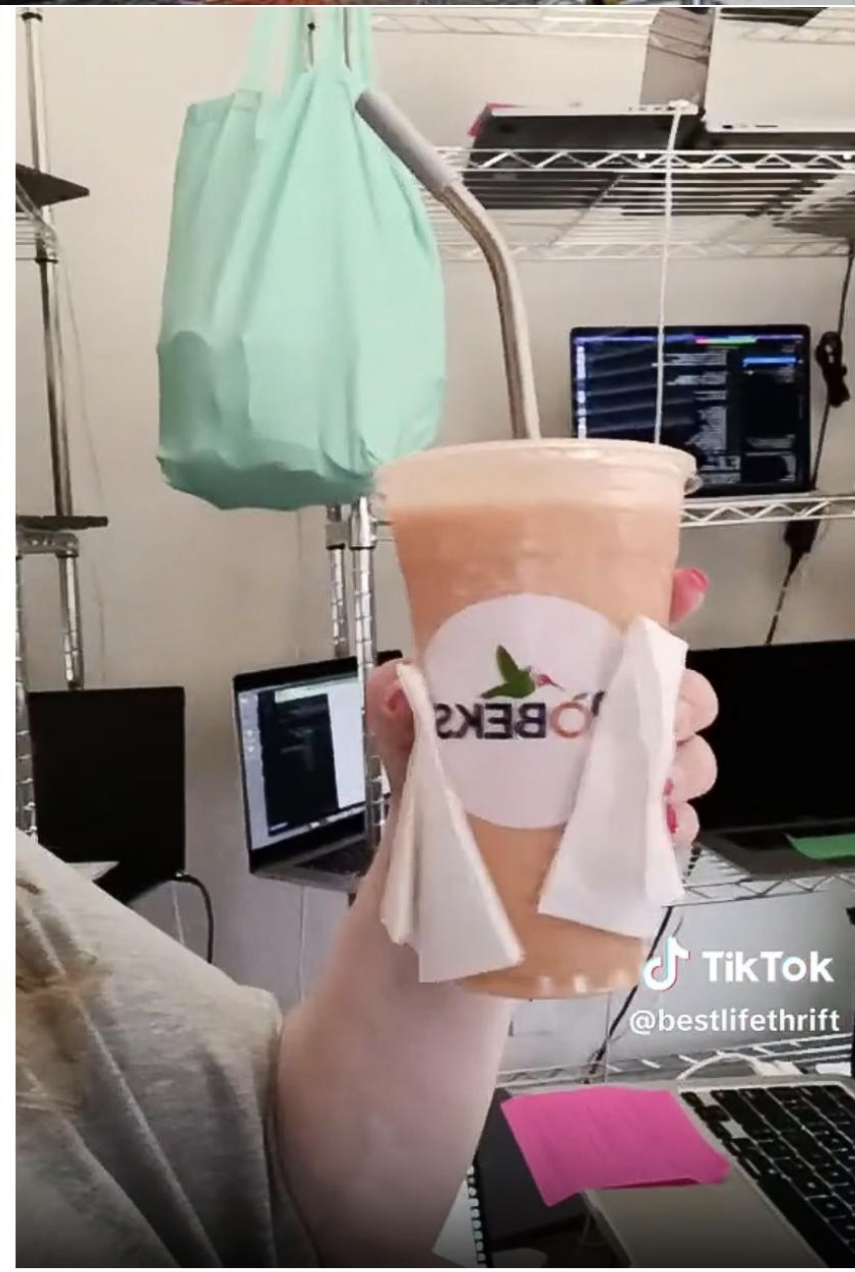
https://www.nationalguard.mil/.../guard-identifies-ai...

Synthetic Identities

Real time Video Deep fake Face Swap – And It Continues…





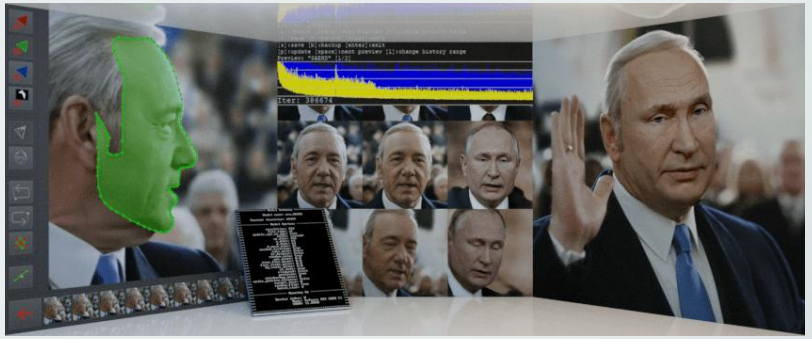


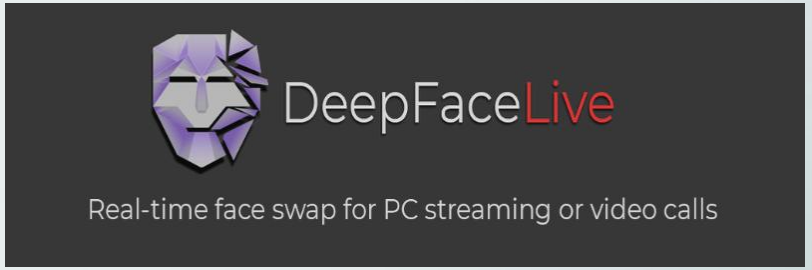
Screenshots from Chapman's June 2023 Tiktok video with laptops in the background.

What Does These Organizations Have in Common?

Create Deepfakes

DEEPFAKE OS





Source: https://deepfakedashboard.com/deepfakeos

Deepfake Creation Process

Determine the mark

Identify the target for the deepfake

Source Videos, Audio

Collect necessary media materials

Upload Audio to Voice Generator

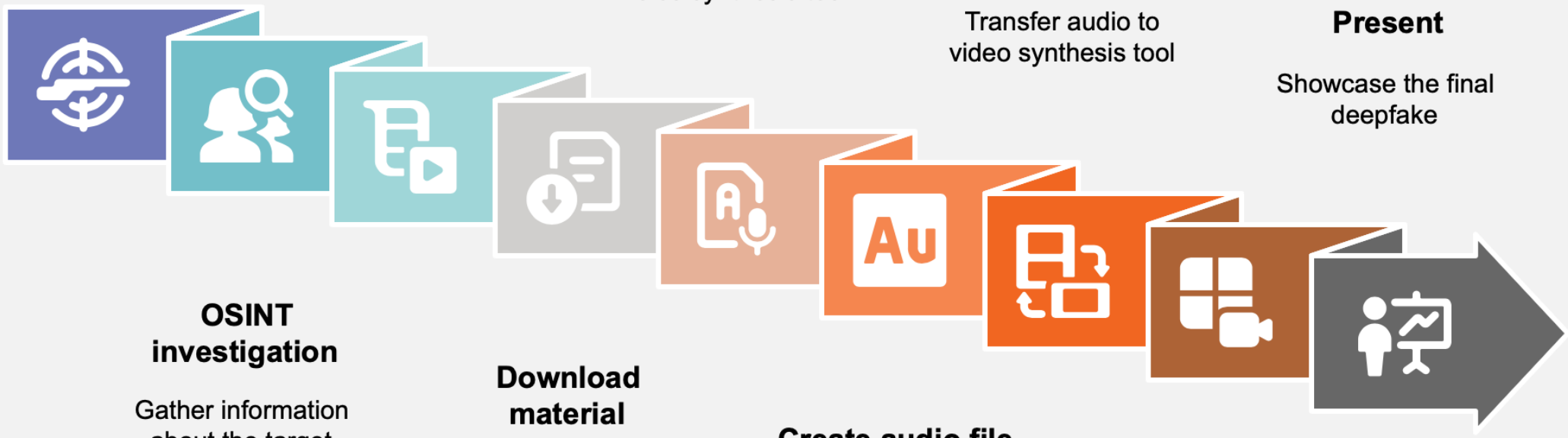
Transfer audio to voice synthesis tool

Upload to Video Generator

Transfer audio to video synthesis tool

Present

Showcase the final deepfake



OSINT investigation

Gather information about the target

Download material

Acquire the media files

Create audio file from text

Generate audio from text input

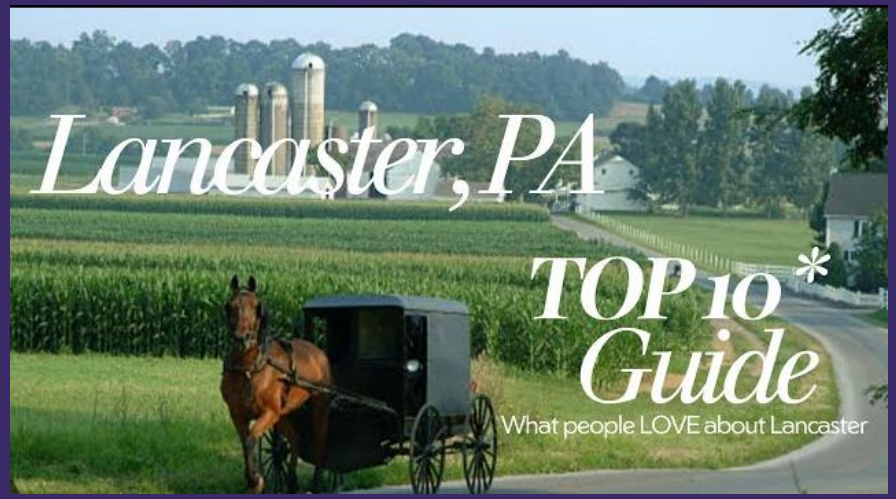
Generate Video

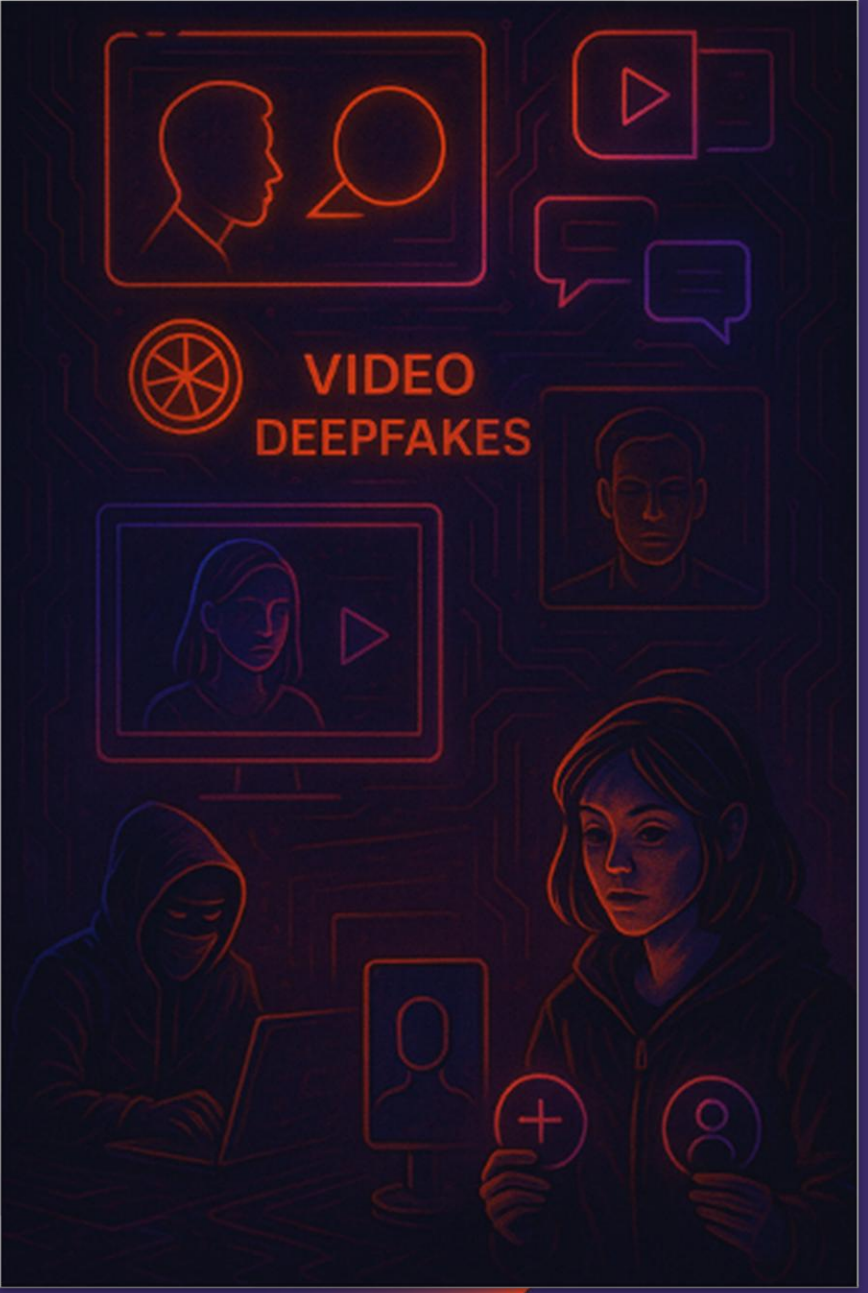
Create the deepfake video

Audio Cloning / Deepfakes

- **Cloning Tools**
  - ElevenLabs, PlayHT, Resemble
- **LLM & Interactive Tools**
  - Dialpad, VAPI
- **Realtime Audio Cloning**
  - Voice.ai, RTVC, Altered.ai
- **Audio Downloaders**
  - YouTube Downloader, Airy, ytmp3.cc
- **Audio Cleaning**
  - Audacity / Audition / Descript

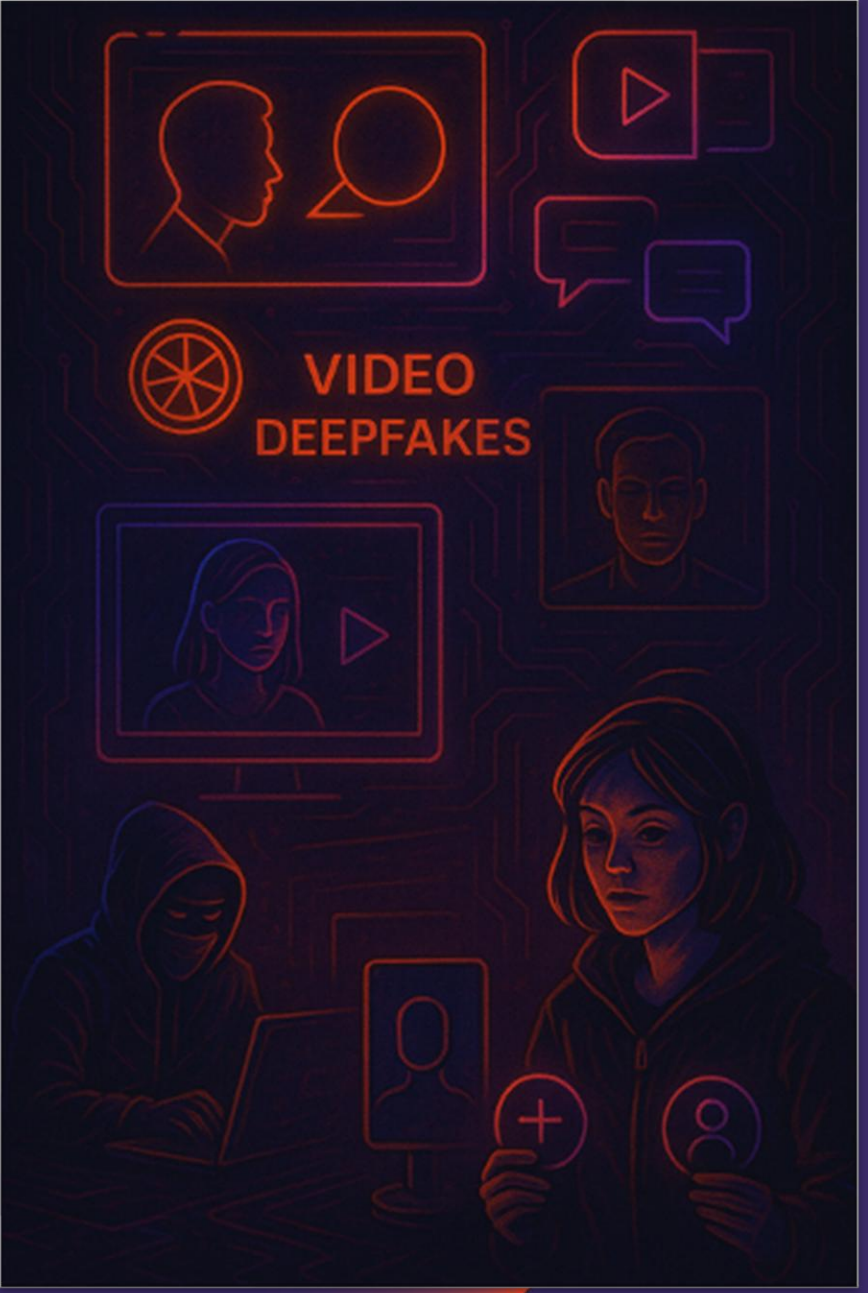
Audio Tips



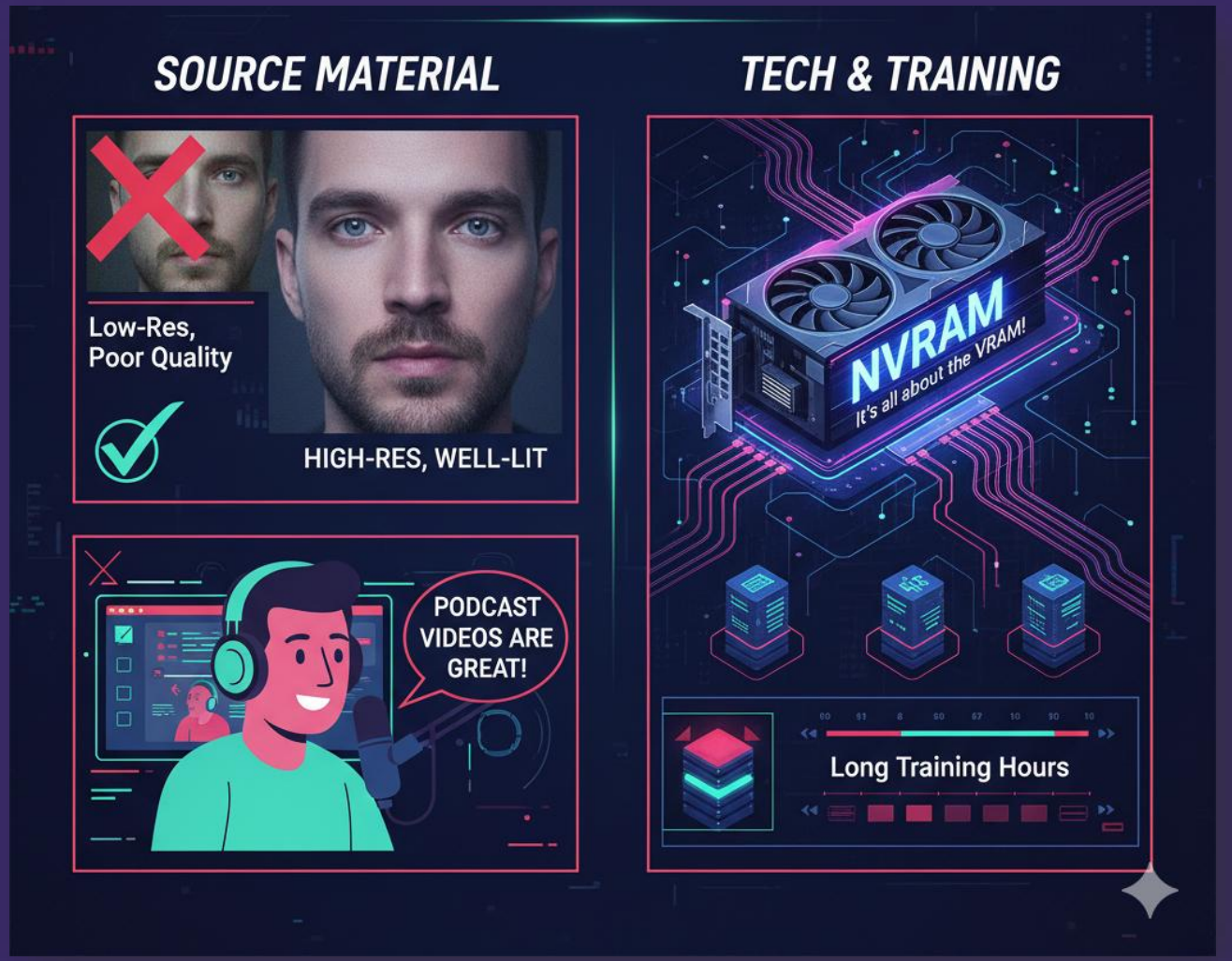


Video Cloning / Deepfakes

- **Image to Video**
  - Hedra, LemonSlice, Synthesia.io, HeyGen, Deepfake Offensive Toolkit (DOT)
- **LipSyncing (Advanced)**
  - KlingAi, Pixverse, Synclabs, Akool
- **Realtime FaceSwap** - Advanced
  - DeepFaceLab, Deep-Live-Cam, SwapFace, MagicCam
- **Video Generation**
  - VEO3, KlingAI, Pixverse



Video Tips



DEMONSTRATION

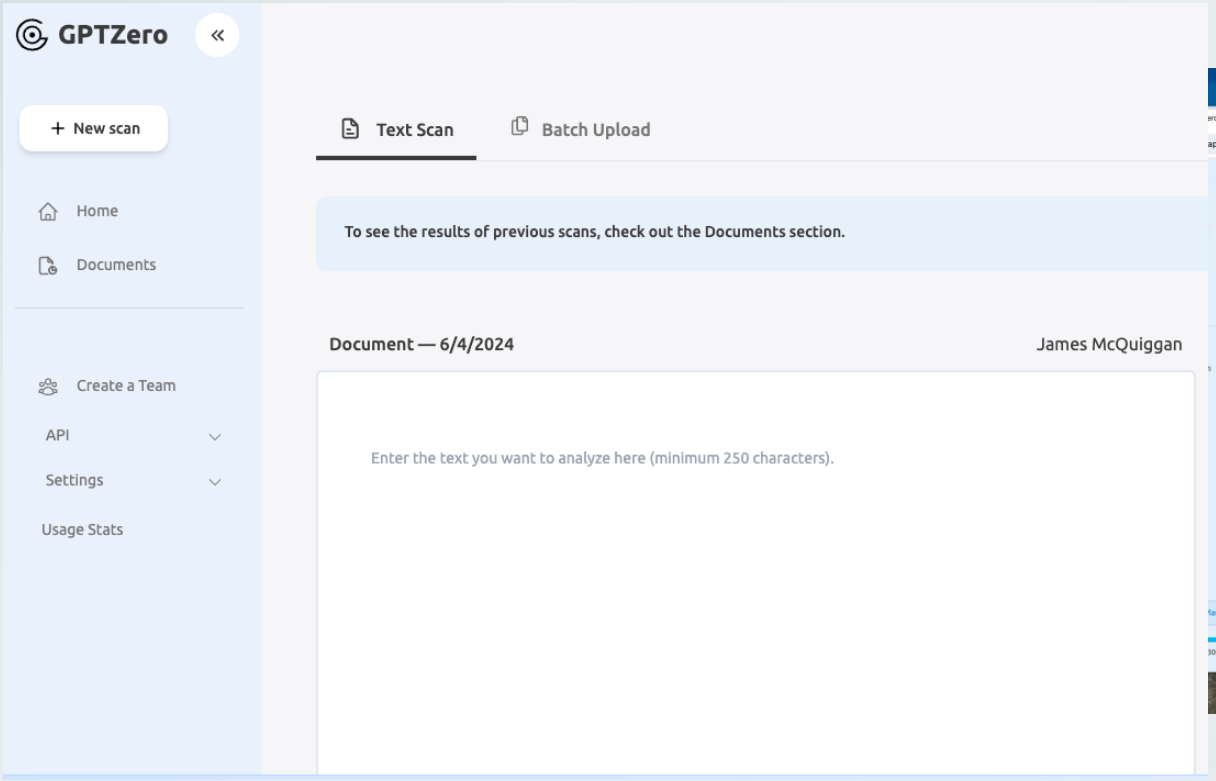
LIVE CISO SACRIFICE OR
COMMODORE 64

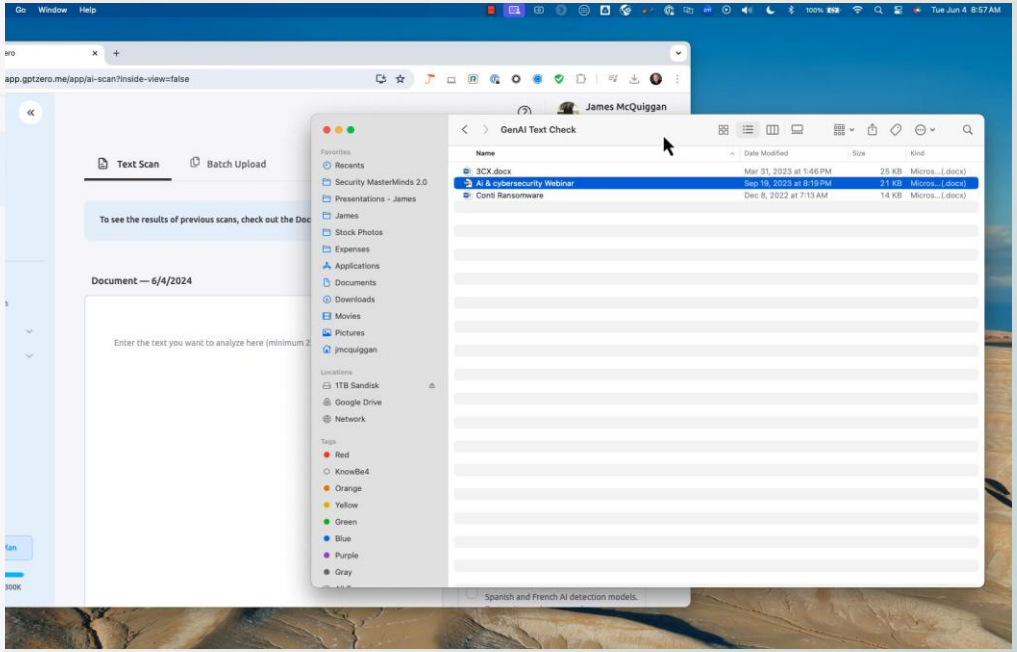
DEMO

Detect Deepfakes

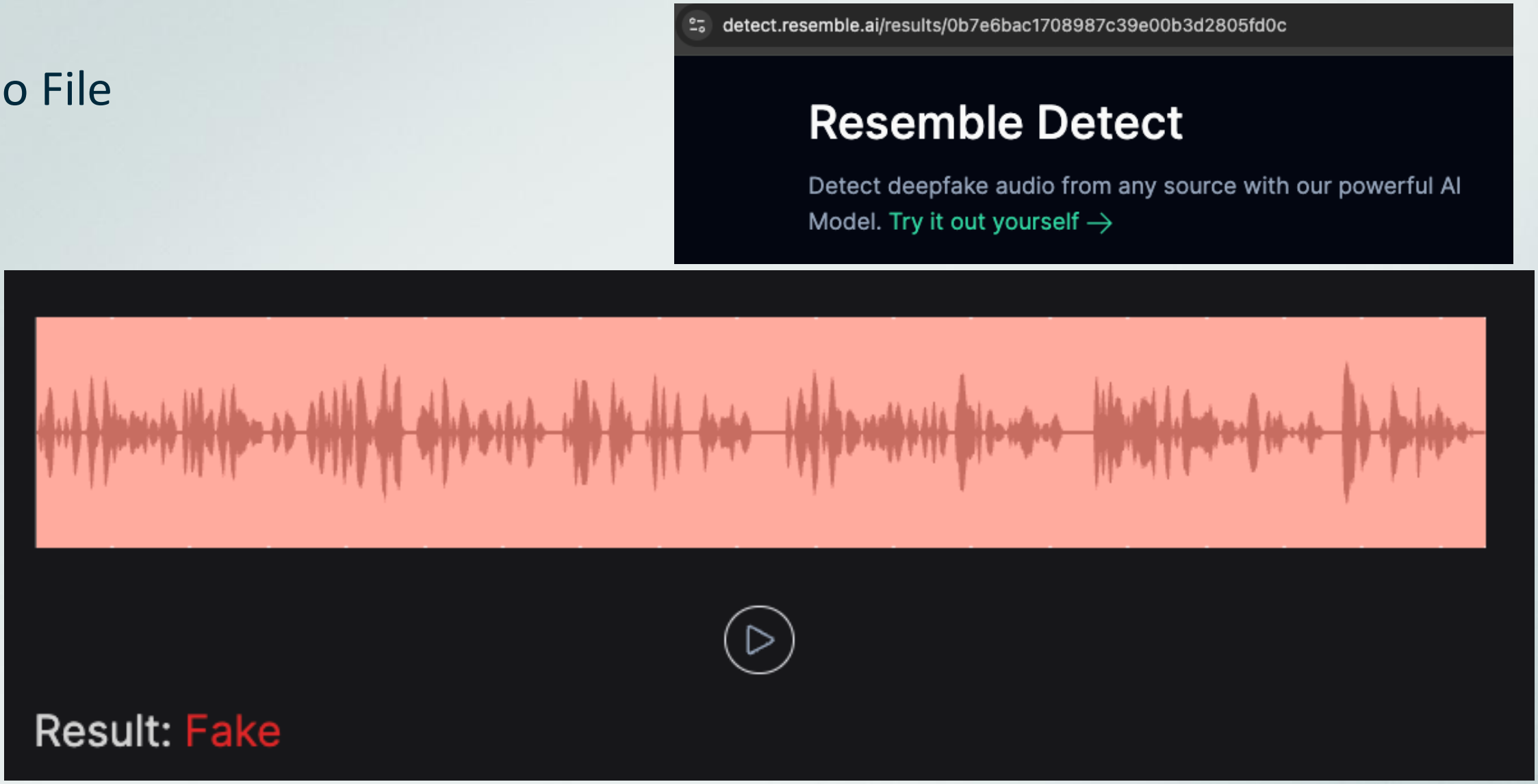
GPTZero – GenAI Synthetic Text Detection

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



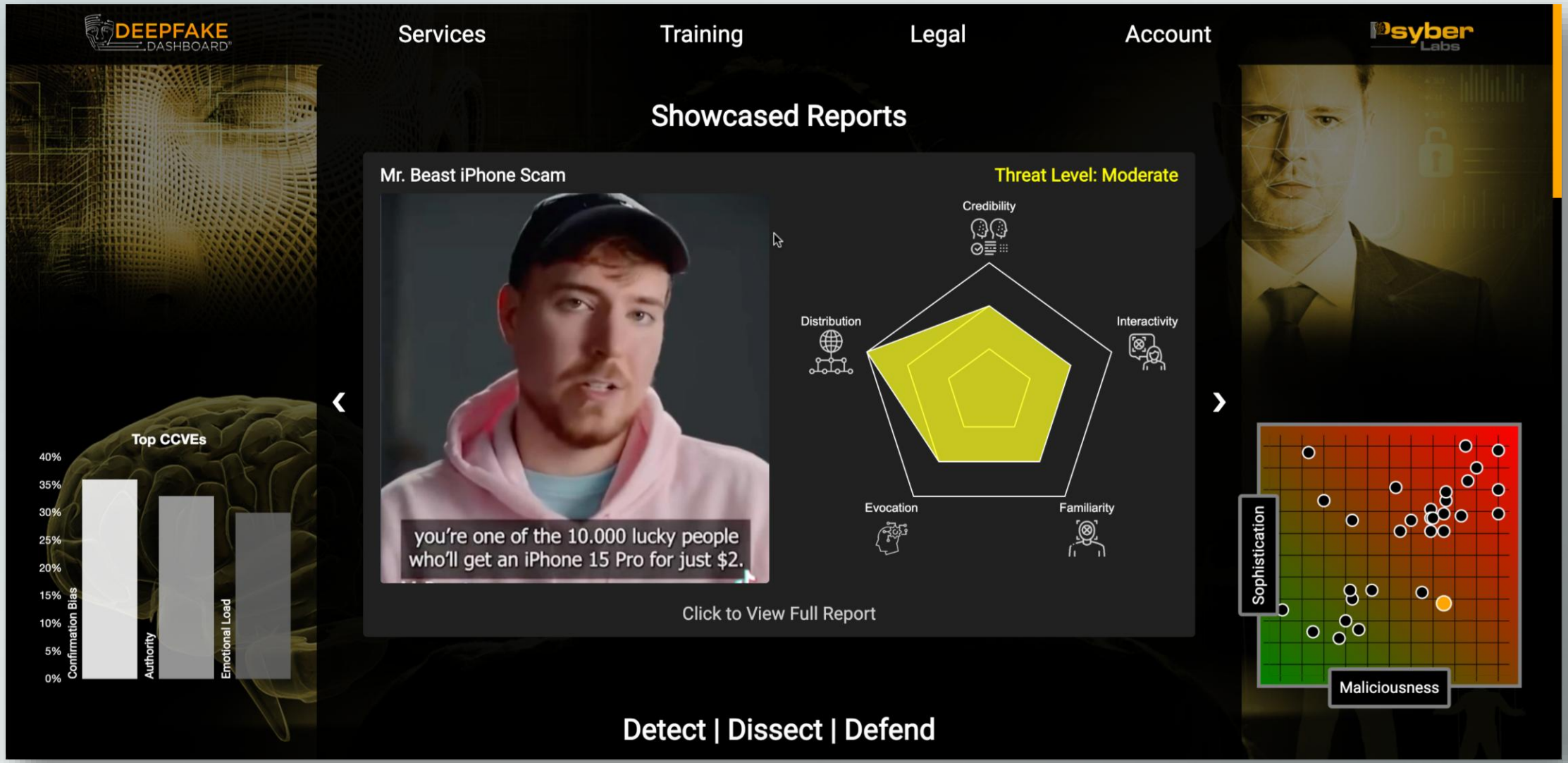


Audio File



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

Deepfake Dashboard



10 Best AI DeepFake Detector Tools

Synthetic Video Detection Challenges

Lack of Standardization

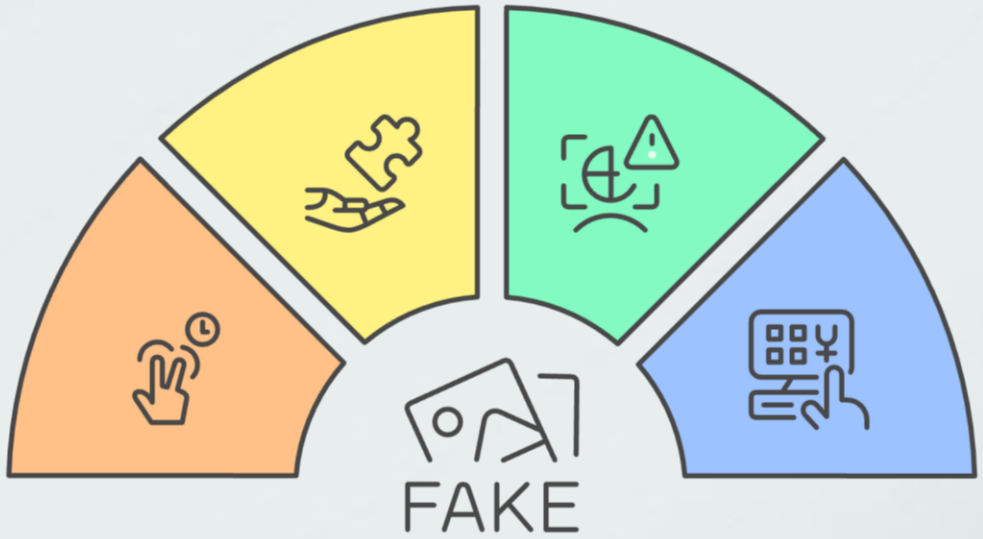
No uniform method exists for detecting deepfakes effectively.

High False Positives

Many legitimate images are incorrectly flagged as deepfakes.

Non-Real-Time Detection

Detection processes do not operate in real-time, delaying responses.

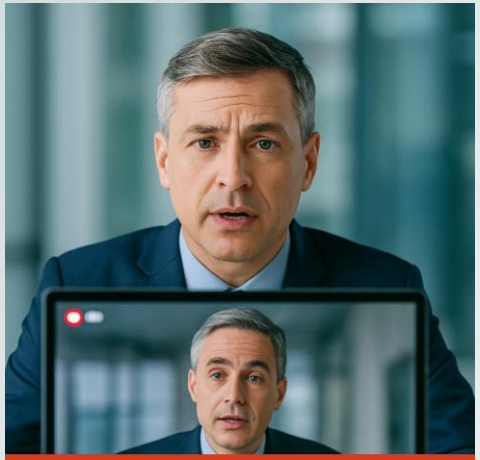


Manual Effort Required

Detection still relies heavily on human intervention and analysis.

Generation technology is outpacing detection technology

Processes / People - Tabletop Exercises



DEEPFAKE CEO WIRE TRANSFER SCAM

A video call appears to come from your CEO while traveling abroad. They urgently ask the CFO to authorize a wire transfer for a confidential acquisition.

EMPLOYEE FALLS FOR DEEPFAKE CEO WIRE TRANSFER SCAM

An employee receives a video call from the supposed CEO, mimicking their appearance and voice. Tricked by the realistic deepfake, they initiate a fraulential

DEEPFAKE PHONE CALL SCAM

You receive a call that appears to be from your manager. Their voice urgently asks for sensitive information to resolve a payroll issue.

FINANCIALLY DISTRESSED BUSINESS SCAM

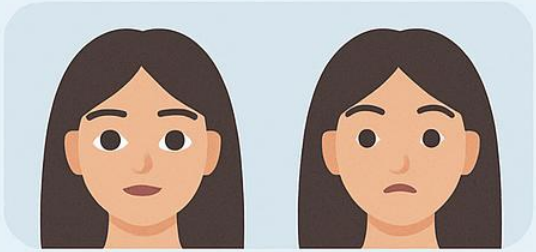
A struggling small business is pressured to make purchases or pay invoices, falsely hoping that it will improve their finances.

PHISHING SCAM

You receive an email claiming to be from your bank. To prevent your account from being locked out, it urgently tells you to click a link and take action.

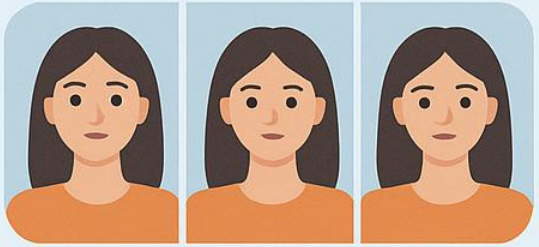
People - Ways to Detect Deepfakes

Look for unnatural blinking



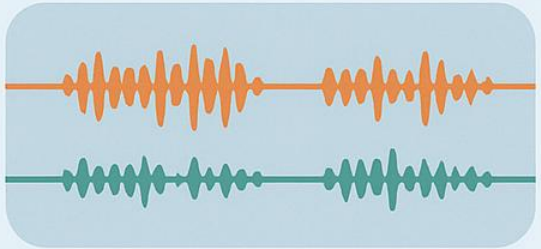
Check lighting consistency

Watch for temporal glitches

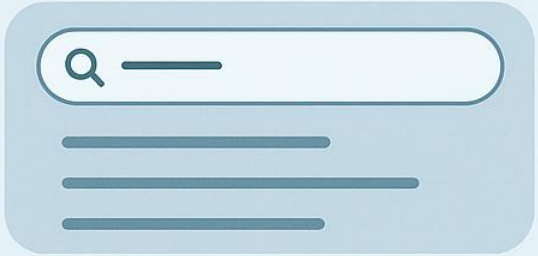


Examine edge artifacts

Listen for audio sync issues



Verify through reverse image search



People - What Should We Be Asking?

✗ Is this a deepfake?

✓ Consider these questions...

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

Using Questions / Be Curious

"What is the book that you recommended to me?"

Defense in Depth for AI & Social Engineering

MFA	• Use MFA Wherever Possible - Non-phishable MFA too • Avoid SMS and verify all requests that you didn't initiate
Email Auth	• DMARC / DKIM / SPF to prevent email spoofing! • Only 15% of orgs use this.
AI vs AI	• Use AI with your cybersecurity gateways & products • Behavioral Analysis in email and users
Zero Trust	• Be mindful of your inbox. Be skeptical, if you're not expecting it and it's a strange or unusual request
Training	• Frequently educate and assess your users • Leverage threat intelligence for your threat landscape

People are a critical layer within security programs

TRUST & VERIFY

BE SKEPTICAL

POLITELY PARANOID

Questions

Final Words

- Wrap-up

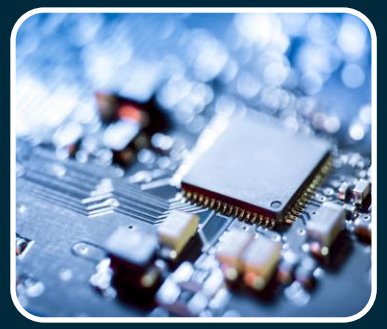
ACTION PLAN

People (Security Awareness & Training)

- Human Risk Management Program
- Conduct deepfake awareness training for all employees
- Implement verification protocols for executive or financial requests
- Use security questions in high-risk communications

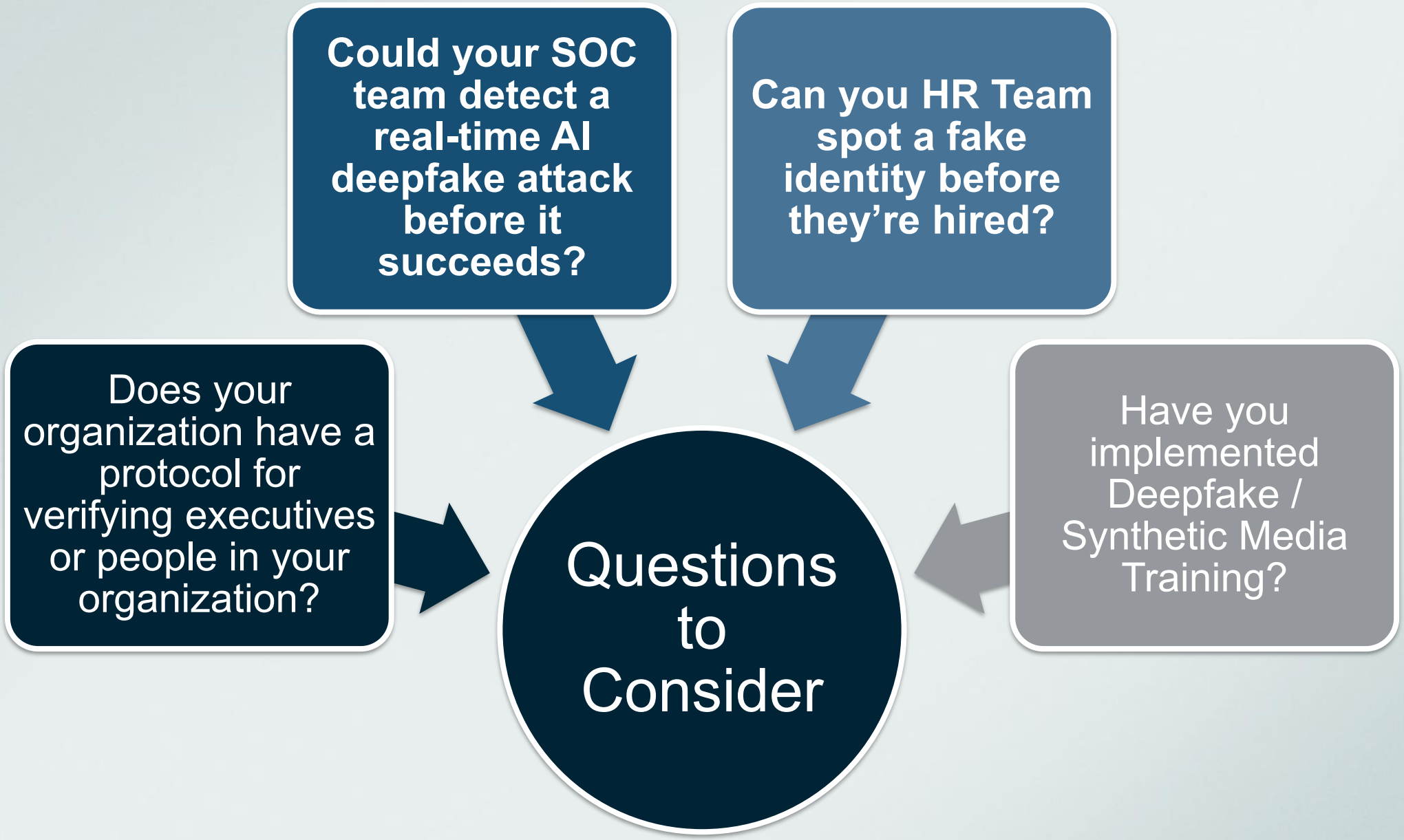
Processes

- Ask location-based verification questions
- Require cameras on for remote interviews with assessments
- Establish a SOC protocol for real-time AI threat detection
- Table Top Exercises Assessing Deepfake Scenarios

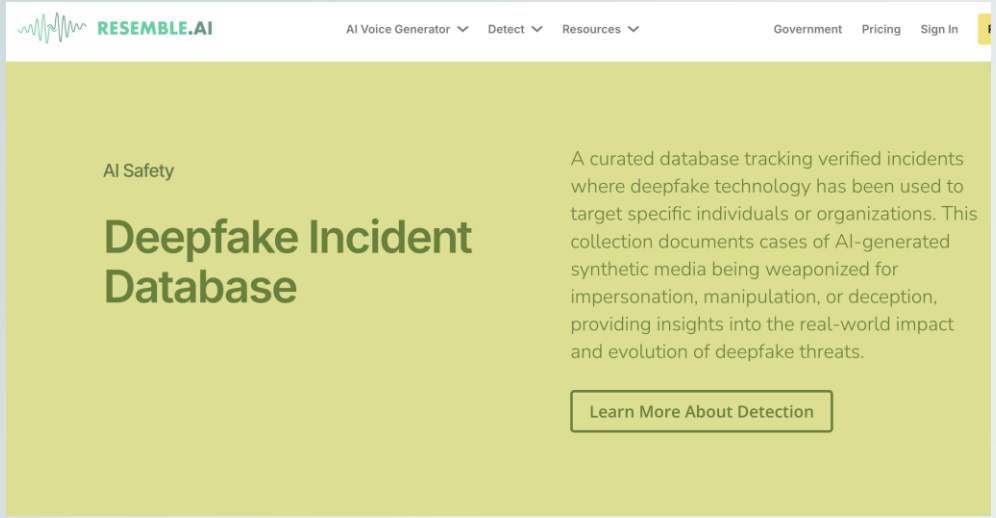


Technology

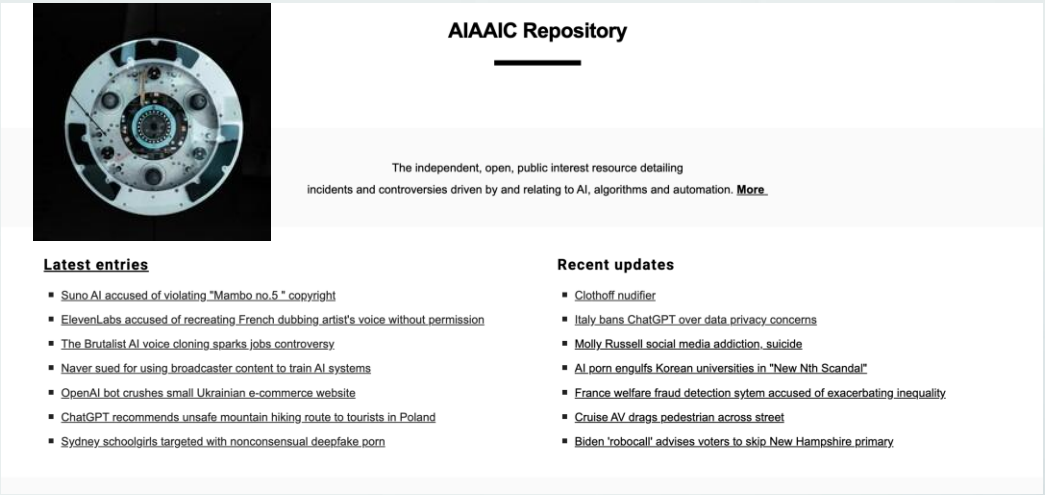
- Deploy AI-driven deepfake detection software
- Implement audio and video authentication measures
- Leverage POCs for deepfake detection services / software



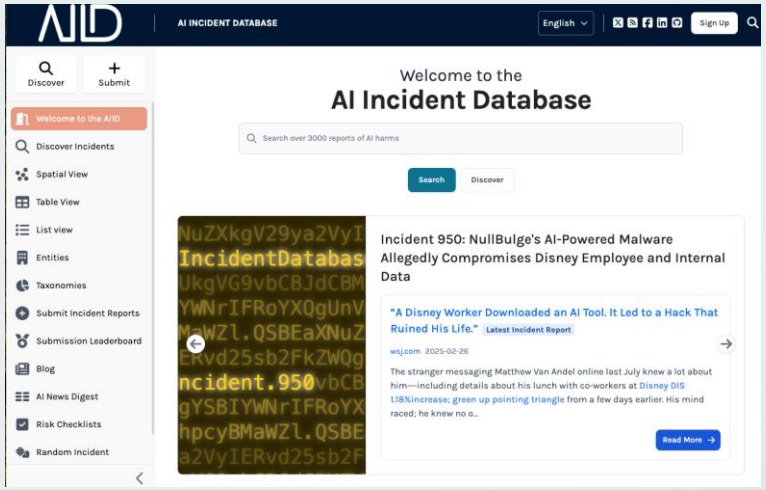
Interested in Learning More?



https://www.resemble.ai/deepfake-database/



https://www.aiaaic.org/aiaaic-repository



https://incidentdatabase.ai

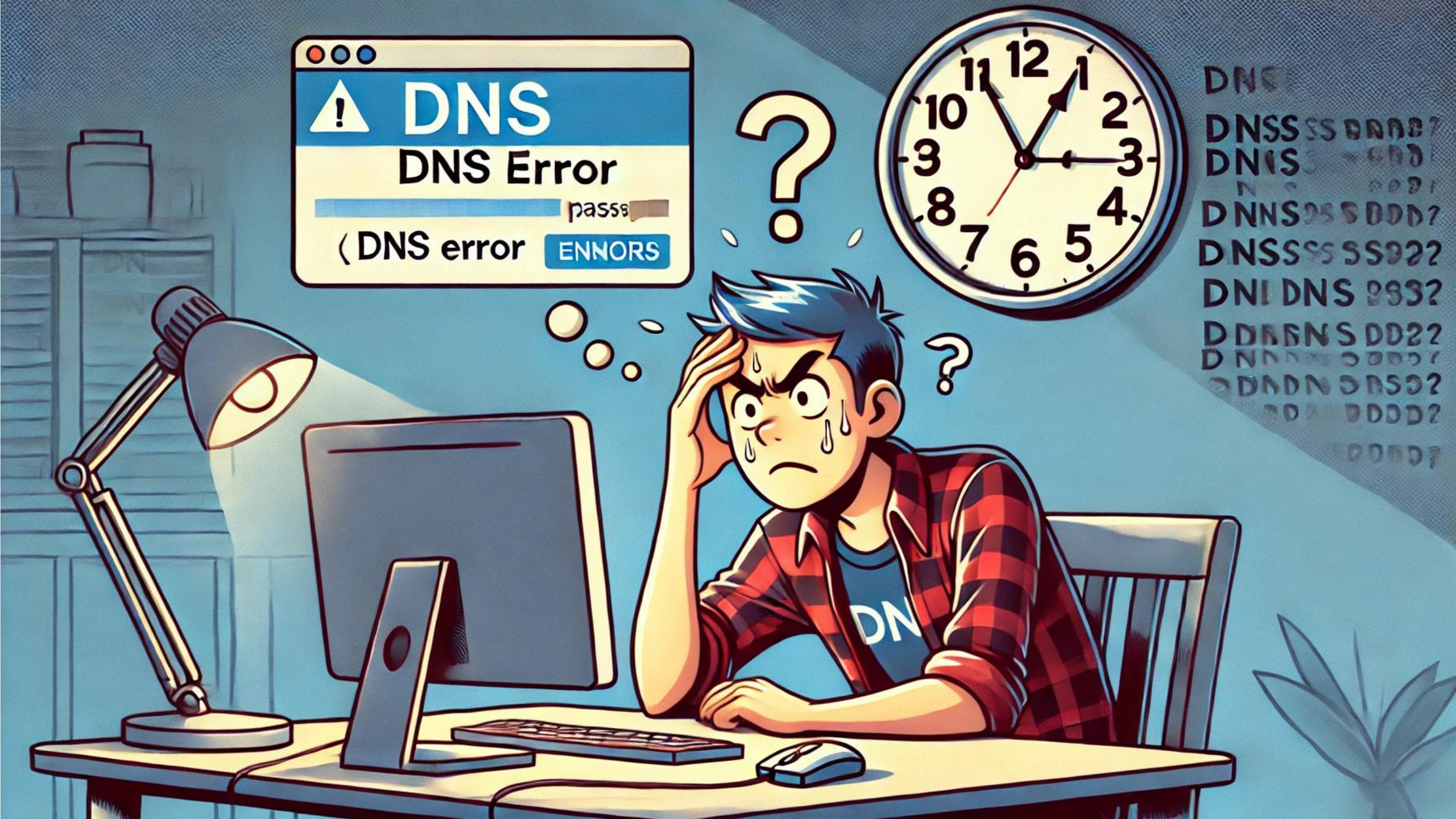
Education, Preparation, and healthy security habits are the only defens

Most Secure Woman?

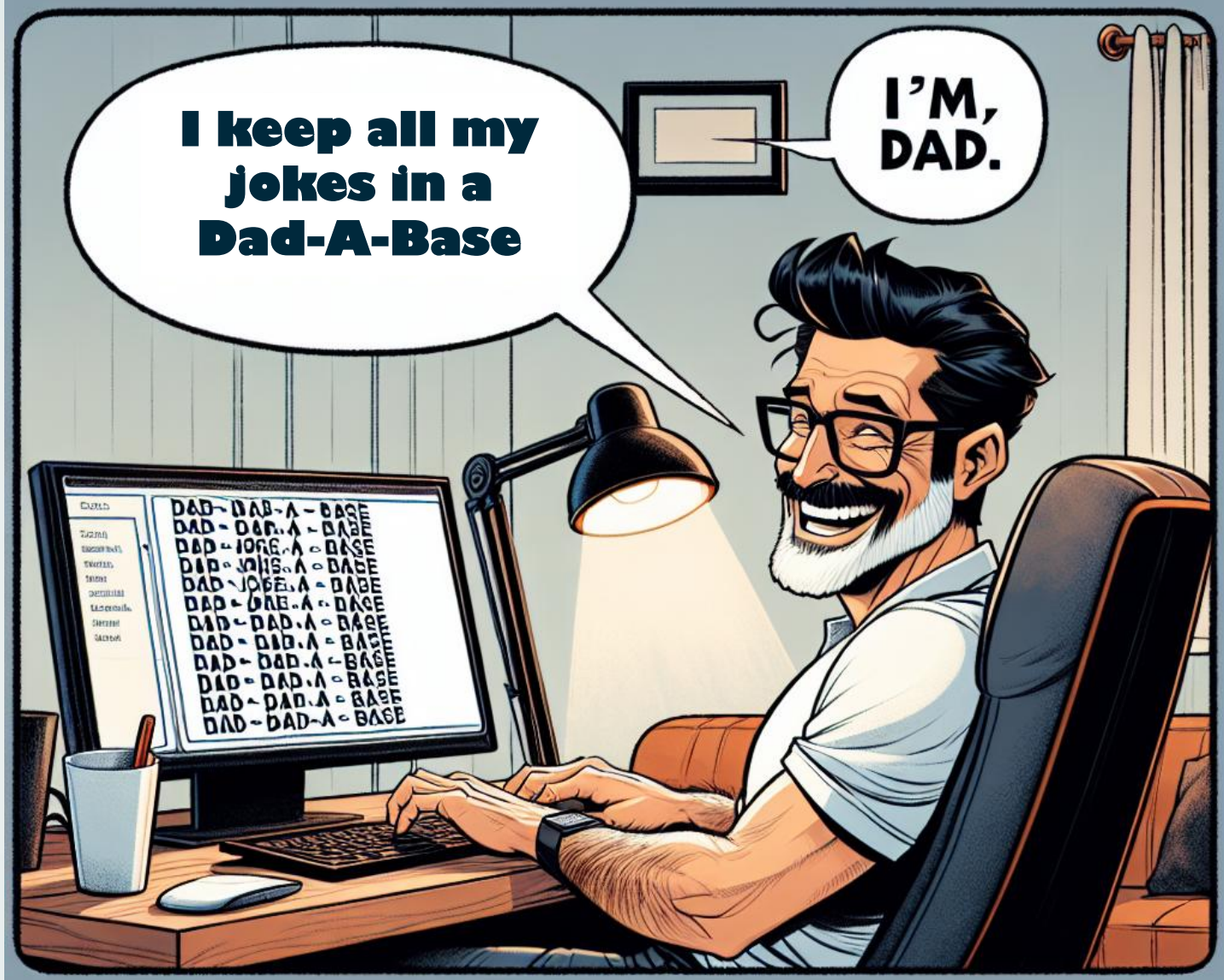
Emma Faye

MFA

Multifactor Authentication


DNS
DNS Error
pass
(DNS error
ENNORS

Yes… I have a way of keeping track of my **Dad Jokes**



I Have a Favor to Ask You

jmcquiggan@knowbe4.com

https://talk.ac/jamesmcquiggan?code=TALK