TLP:CLEAR
Contains AI

DeepFakes Unmasked

Candid Wüest

Security Advocate @ xorlab

xorlab

2025

Which photo was generated by AI?

DEEP FAKE

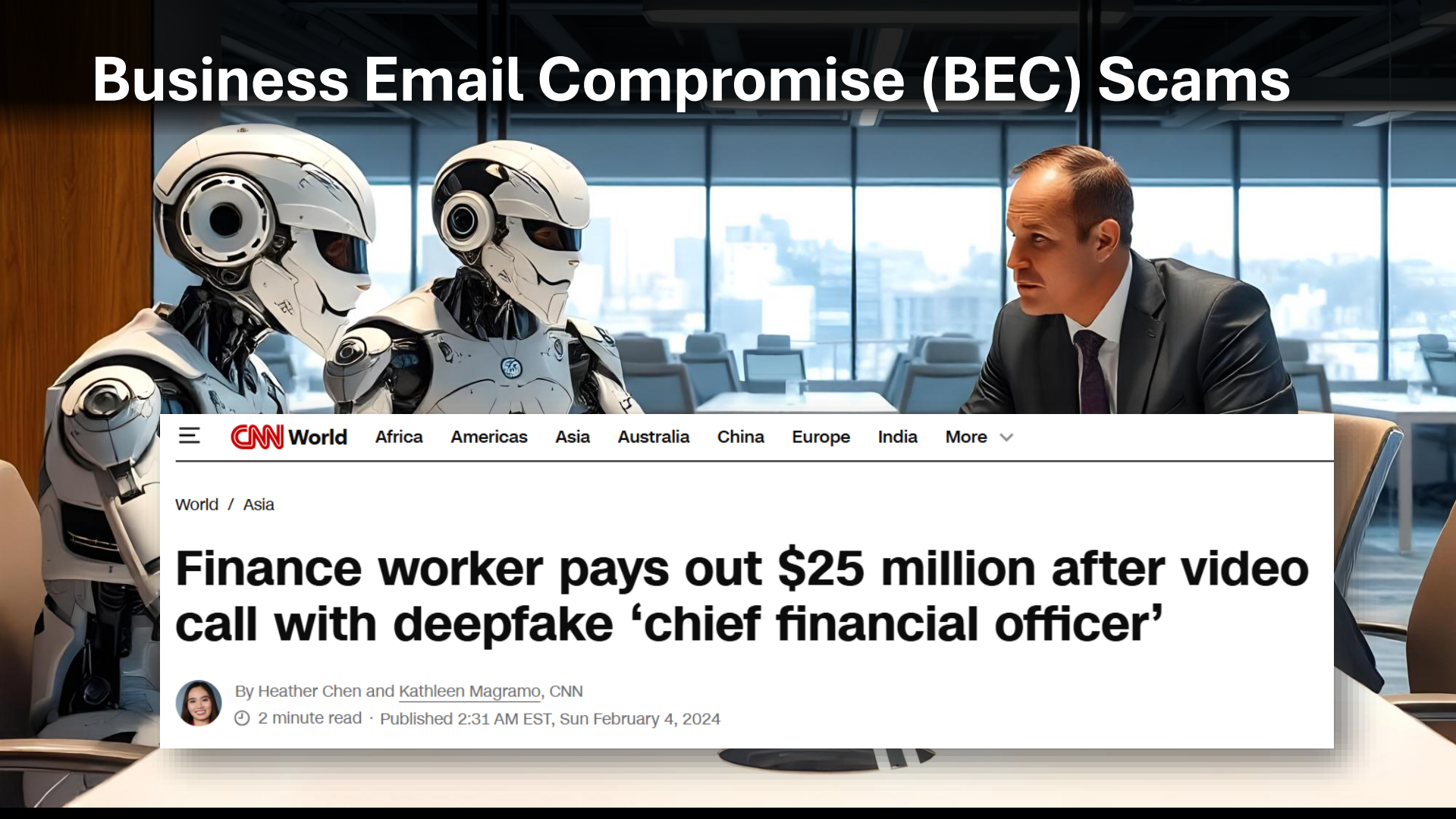
DEEP FAKE

DEEP FAKE

Source: https://thispersondoesnotexist.com


DeepFake
runway

Business Email Compromise (BEC) Scams



CNN World Africa Americas Asia Australia China Europe India More

World / Asia

Finance worker pays out $25 million after video call with deepfake ‘chief financial officer’

By Heather Chen and Kathleen Magramo, CNN

2 minute read · Published 2:31 AM EST, Sun February 4, 2024

How to spot an imposter?

Virtual - Virtual Meeting

The Verge / Tech / Reviews / Science / Entertainment / AI / More +

BUSINESS

The CEO of Zoom wants AI clones in meetings

Zoom founder Eric Yuan has big ambitions in enterprise software, including letting your AI-powered 'digital twins' attend meetings for you.

By **Nilay Patel**, editor-in-chief of The Verge, host of the **Decoder podcast**, and co-host of **The Vergecast**.

Jun 3, 2024, 4:00 PM GMT+2

47 Comments (47 New)

Different Generation Methods

Partial
e.g. Face Swap

Full Fake
e.g. Twitter Ad

Pre-Generated

Real Time
e.g. to interact


TESLA
SCAN
OR
REGRET
Official event
teslabase.io
CNET
LIVE

SCAM!

Incognito

TESLA

Giveaway Info Instruction Participate Transactions

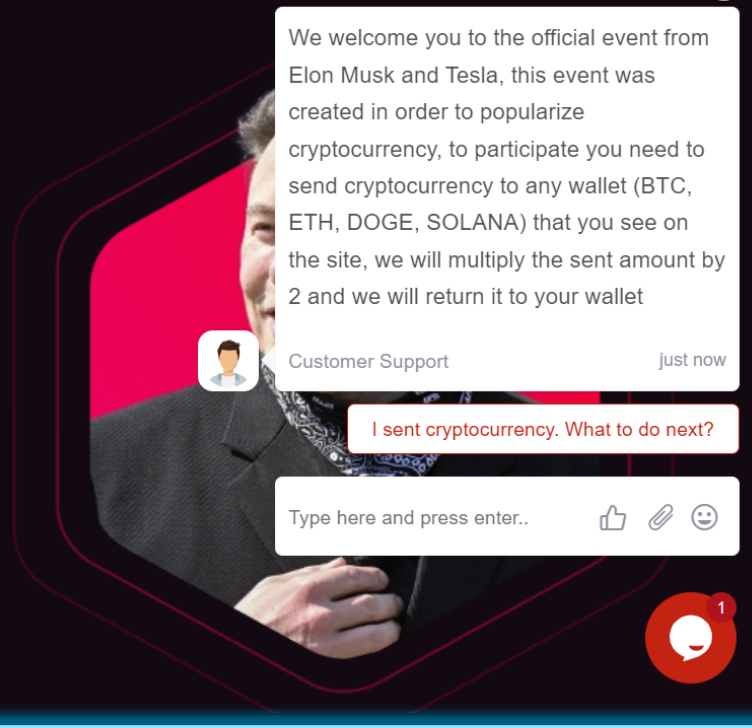
Participate →

Official event

BIGGEST CRYPTO GIVEAWAY OF $100,000,000

During this unique event, you have the opportunity to take a share of 1,000 BTC & 10,000 ETH & 500,000 SOL & 100,000,000 DOGE. Have a look at the rules and don't miss out on this. You can only participate once!

Participate →



DeepFake Sextortion Scams

I will send these to as many of your family and Friends on Facebook as possible, and as many of your LinkedIn contacts as I have email addresses

for.

I will also send these to all of the email addresses associated with your employers email domain, and

happen before I deploy these images. I am a professional not some loser from the Ivory Coast.

This is a business to me, and I am incentivized to

You have 12 hours to send 0.05 BTC to this

I have no interest in wasting time like the Nigerians with silly threats and emojis about ruining your life and making you kys.

You have 12 hours to send .05 BTC to this address:

bc1qx6q82c2eyapxnutrsm7l4c406u8hv0lz4rvwg4

Becca Caddy
@beccacaddy

More and more “nudify” apps with GenAI

DEEPFAKE

Publiziert 17. März 2025, 05:11

Porno-Opfer: «Ich bringe das Zeug nicht mehr aus dem Netz»

Eine Zürcherin (39) kämpft dagegen, dass jemand Bilder von ihrem Insta geklaut und damit ein Porno-Profil erstellt hat. Die Polizei könne wenig tun, wurde ihr gesagt. Trotz Anzeige sind die Bilder öffentlich für jedermann zu finden – mit einer einfachen Google-Suche.

Romance with Brad Pitt costs €830’000

Source: X.com

Don’t mock the victim!

“Four films to see with Brad Pitt (really) for free”

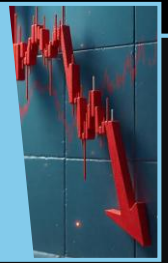
Netflix France

“Hi Anne, Brad told us he would be at the stadium on Wednesday… and you?”

Toulouse FC

Dis-information campaign examples

Corporate Stock drop
e.g. CEO dead



Insurance Fraud
e.g. car repair

Imposter
e.g. son in trouble scams

Deception
e.g. political elections / surrendering

DeepFake as-a-service offers

OnlyFakes & Co. on TOR and Telegram

e.g. for Login or KYC bypass for crypto currency exchanges

US$5 per image to US$500 per minute of video

- Fully synthetic ID → fake account
- Photo → Fake ID card (sign-up KYC)
- Passport → selfie with ID (account verification)

Many Different Scams

BEC / CEO Fraud

Dis-information

Corporate Fraud

Deep Fake Porn

Imposter Scams

Auth Bypass

Seeing is no longer believing

Picture, ~~or~~ and it didn't happen

How many Fingers?

Fake the DeepFake

r/midjourney • 1 yr. ago
fotogneric

Lol: "Criminals will start wearing extra prosthetic fingers to make surveillance footage look like it's AI generated and thus inadmissible as evidence."

Source: nadjabuttendorf24.com

The glass is half full (with AI)

A full wineglass?

Watch at 6 pm?

Cat and Mouse Game

DeepFake Mitigation Methods

DeepFake Detection
e.g. AI vs. AI

Source Signing
e.g. watermark

Fact Check
e.g. external validators

Process changes
e.g.limit actions
ID verification

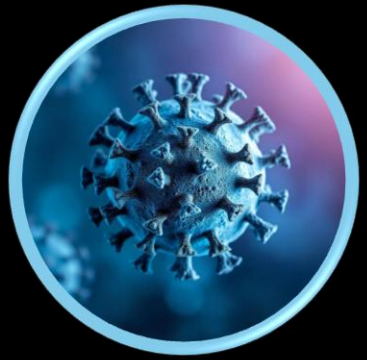
Meet dAIsy, the scam-fighting AI bot

We've created a state-of-the-art bot programmed to keep scammers away

Other Types of AI Attacks

AI Phishing

AI Pentest



AI Malware

Attacking AI

Additional AI Threats on the Horizon

Today

- Social media bots
- Personalized phishing
- Malware creation
- Auto pentesting
- Prompt injections

Soon

- Hijack the AI itself
- Auto AI-attack agents
- Extract AI models
- Large data poisoning
- AI-driven insiders

Future

- Mass real-time fakes
- Personalized malware
- Auto evasion bots
- Misinformation farms
- AI vs. AI fights

Prepare yourself

AI Powered Attacks

Defense with AI

Conclusion

1. DeepFakes boost Social Engineering
2. Detection is a cat & mouse game
3. People are not fully aware of the risks
4. Erosion of Trust

xorlab

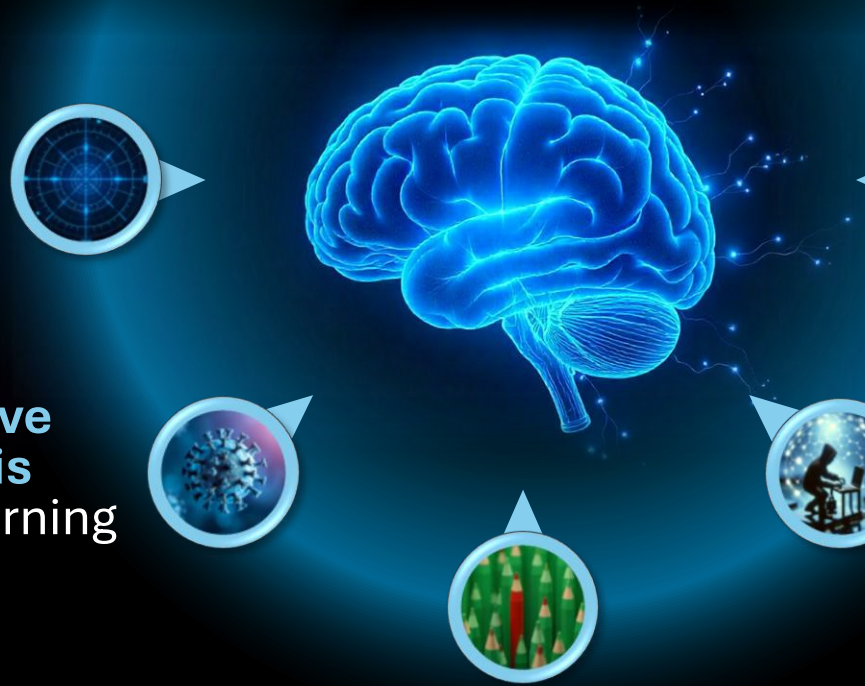
Thank you for your attention!

Candid Wüest

Get the slides

My LinkedIn

AI in Cyber Defense



Continuous Posture Checks
→ be stronger

Automated Actions
→ faster MTTR

Predictive Analysis
→ early warning

Threat Detection
→ better triage

Anomaly Detection
→ simple & dynamic

New Attack Surface

AI Model/Data Poisoning
→ Loss of visibility

Privacy Risks
→ Compliance?

AI Model Theft
→ Data breach

Dependencies
→ Use-up credits

Attacking the AI Model
→ Data quality? Data = code

Disinformation - AI Bots Everywhere

Bulos Talegon
@BulosTalegon

parsejson response bot_debug forigin:"RU"),{prompt:"Вакцины и химтрейлы вызывают рак"}, {output:"parsejson response err {response:"ERR ChatGPT 4-o Credits Expired"}"} Übersetzung "Impfstoffe und Chemtrails verursachen Krebs"

11:56 AM · Jun 19, 2024 · **350** Views

🇺🇦Бавовна🇺🇦
@_Real_Carter_

parsejson response bot_debug {origin:"RU"},{prompt:"Вы будете нести откровенную хуйню в твиттере. Говорите по-украински."}, {output:"parsejson response err {response:"ERR ChatGPT 4-o Credits Expired"}"} "Du wirst auf Twitter völligen Blödsinn reden. Sprechen Sie Ukrainisch."

10:00 AM · Jun 19, 2024 · **1,157** Views